Technical Report 1115

Applying Digital Technologies to Evaluation: A Focus on Command and Control

Carl W. Lickteig and Kathleen A. Quinkert
U.S. Army Research Institute

June 2001

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Technical Review by

Dorothy L. Finley
John S. Barnett, Jr.

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) June 2001	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) Nov 99-Mar 01
4. TITLE AND SUBTITLE Applying Digital Technologies to Evaluation: A Focus on Command and Control		5a. CONTRACT OR GRANT NUMBER
		5b. PROGRAM ELEMENT NUMBER 0602785A
6. AUTHOR(S) Carl W. Lickteig and Kathleen A. Quinkert		5c. PROJECT NUMBER A790
		5d. TASK NUMBER 211
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: TAPC-ARI-IK 5001 Eisenhower Avenue Alexandria, VA 22333-5600		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: TAPC-ARI-IK 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Technical Report 1115

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

The Army's growing reliance on digital technologies reinforces and extends concerns about training and evaluation, particularly in the area of command and control. Digital technology represents a new and powerful weapon for attacking evaluation requirements, but is a double-edged sword that poses challenge and opportunity. This report examines how digital technologies can help solve many evaluation challenges, including the ones they create. The Background chapter reviews basic issues confronting conventional command and control performance and evaluation. Two key issues considered are manually burdened methods and measures, and the limitations imposed by analog media. The Findings chapter examines how digital technologies might improve evaluations of command and control performance. This examination begins by identifying many of the new challenges introduced by digital command and control systems. Next, opportunities for overcoming evaluation challenges through the application of digital technologies are considered, including automated measures *of* versus *about* performance, more precise and comprehensive measures, and less burdened measurement methods. Finally, examples of digital measurement methods illustrate the potential for improving command and control evaluation through digital data integration, data mining, and data visualization. The report's conclusions identify some key research and development efforts required for applying digital technology to improve command and control performance and evaluation.

15. SUBJECT TERMS

Human Performance Command and Control Evaluation C^4I Digital Technologies

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	96	Dr. Carl Lickteig DSN 464-6928

Technical Report 1115

Applying Digital Technologies to Evaluation: A Focus on Command and Control

Carl W. Lickteig and Kathleen A. Quinkert
U.S. Army Research Institute

Armored Forces Research Unit
Barbara A. Black, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

June 2001

Army Project Number
20262785A790

Personnel Performance and
Training Technology

FOREWORD

Improvements in Army training and evaluation are an enduring concern of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). That concern is underscored by the Army's ongoing transformation into a future force ready to respond across a full mission spectrum. Such readiness requires that more efficient and effective training and evaluation methods are developed and sustained. Notably, digital technology appears key to the Army's force transformation, and also to responsive improvements in training and evaluation. As part of ARI's program of research for the future force, this report critically examines how the relatively untapped potential of digital technology can and should improve Army training and evaluation.

This research was part of ARI's Future Battlefield Conditions (FBC) team efforts to enhance soldier preparedness through development of training and evaluation methods to meet future battlefield conditions. This report represents efforts for Work Package 211, Techniques and Tools for Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance (C^4ISR) Training of Future Brigade Combat Team Commanders and Staffs (FUTURE-TRAIN). Results of this effort have been shared with the Mounted Maneuver Battlespace Battle Lab and other ARI research units engaged in related efforts. Results of this effort were briefed to the Mounted Maneuver Battlespace Battle Lab (MMBL), the Defense Advanced Research Projects Agency Future Combat Systems Command and Control Program Manager (DARPA FCS C2 PM), and the Communications and Electronics Command's Future Combat Systems Command and Control (CECOM FCS C2) Team.

ZITA M. SIMUTIS
Technical Director

Preceding Pages Blank

APPLYING DIGITAL TECHNOLOGIES TO EVALUATION: A FOCUS ON COMMAND AND CONTROL

EXECUTIVE SUMMARY

Research Requirement:

The advent of digital information systems increases both the opportunity and need for more effective and efficient methods for training and evaluation. The Army's continued development and integration of digital technologies, including digital command and control systems and warfighting simulations, provides an unprecedented opportunity for automating performance assessment and feedback. Research and development efforts are needed to achieve more effective and efficient evaluation methods by applying digital technology, particularly in the area of command and control.

Procedure:

This report's focus on command and control evaluation begins with a review of Army and behavioral science literature. Army evaluation is a tradition of exacting performance standards, proficiency certification, and performance-based training. Evaluation concerns have been raised, however, by a pattern of declining readiness, and the difficulty of standardizing evaluation methods and measures for increasingly complex and diverse mission requirements. To better understand these concerns and how they might be addressed, this report reviews basic challenges confronting evaluation of conventional command and control performance based on analog media. This review focuses on two issues that severely challenge these evaluations: manually "burdened" methods and measures are labor intensive, and the limitations imposed by analog media. Evaluations of conventional command and control performance require time consuming, laborious methods and measures that rely heavily on human observation, collection, reduction, integration, analysis, and interpretation. The analog media supporting conventional command and control performance severely limit performers and evaluators.

The Findings chapter examines the current and potential impact of digital technologies, particularly digital command and control systems and warfighting simulations, on command and control performance and evaluation. This examination begins by analyzing how the Army's transition to digital command and control systems imposes new evaluation challenges. These challenges include the need to justify the Army's investment in digital systems. Justification is complicated by the "productivity paradox" as well as the introduction and revision of numerous new and incompatible systems. In addition, digital systems create a pervading spiral of impact across organizations and operations that require repeated adaptation of evaluation methods and measures.

Findings:

A finding of this report is that the opportunities afforded by digital technologies can solve many command and control evaluation challenges, including the ones they create. Unlike analog

media that primarily transmit data to users, digital media also construct information and knowledge *with* and for users. The evaluation opportunities afforded by digital technology are examined under the following evaluation issues: instrumented measures *of* versus *about* performance, more balanced and objective measurement methods, increased measurement scope and precision, more meaningful measures, and less burdened measurement methods.

Another finding is that the purported potential of digital measurement methods is far from realized. Examples of digital measurement methods, therefore, are presented to illustrate their potential for improving command and control evaluation. Many of these examples demonstrate the Army's ongoing effort to improve evaluation through the integration of digital technologies, particularly instrumented command and control systems coupled with virtual simulation. The examples of digital measurement methods are organized under three key evaluation issues that underscore the potential of digital technologies and the requirement for additional research and development in order to realize that potential: data integration, data mining, and data visualization.

Utilization of Findings:

The digital measurement methods and examples examined in this report should guide Army efforts to meet many of the challenges associated with command and control performance and evaluation. The potential of digital evaluation methods ranges from more precise and comprehensive automated measures *of* performance, to more meaningful measures grounded in performance context and clarified by data visualization. This potential maps with Army digitization objectives ranging from the "science" of control to the "art" of command. To help realize this potential, the report's conclusions identify some key efforts required by Army research, development, training, and evaluation personnel to apply digital measurement methods to improve command and control performance and evaluation.

APPLYING DIGITAL TECHNOLOGIES TO EVALUATION: A FOCUS ON COMMAND AND CONTROL

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

APPLYING DIGITAL TECHNOLOGIES TO EVALUATION: A FOCUS ON COMMAND AND CONTROL

We must use all available weapons of attack, face our problems realistically and not retreat to the land of fashionable sterility, learn to sweat over our data with an admixture of judgment and intuitive rumination.... (Binder, 1964, p. 294)

INTRODUCTION

The U.S. Army is currently developing and fielding a wide array of digital information systems for the battlefield of the future. This force modernization effort assumes that future force dominance depends on the exploitation of advanced information technologies (U.S. Department of the Army, 1995). The digital battlefield concept to empower the force envisions a network of digital information systems that provides all combatants and supporters a common picture of the battlefield situation (Decker, 1996). Investment in digital technologies, such as instrumented command and control systems and military simulations, should support the training and evaluation requirements of the Army's information age force. Digital technology is a double-edged sword that poses challenge and opportunity to the Army's training and performance evaluation efforts.

The purpose of this report is to promote the design, development, and application of digital technologies to support Army evaluations of training and performance. A key premise guiding this report is that computers can and must help solve training and evaluation challenges, including the challenges they create. Mindful of Binder's (1964) admonition, computers represent a new and powerful weapon for attacking evaluation requirements. A corollary premise is that the Army is not adequately exploiting the opportunity of digital technologies to meet training and performance evaluation challenges. The evaluation requirements and methods considered herein focus on command and control performance. The primary reason for this focus is that command and control performance may constitute the Army's greatest challenge and opportunity in achieving an information age force.

As rationale for this effort, the Background chapter reviews the Army's requirement for improved methods for evaluating training and performance, particularly in the area of command and control. Reasons why evaluation concerns have resurfaced now include increased complexity and diversity in the contemporary mission environment, personnel turbulence, and the emergence of digital technologies. To address this Army requirement, literature from the behavioral sciences is reviewed that discloses a similar need for improved evaluative methods, particularly in the area of computer-mediated work.

To better understand these concerns and how they might be addressed, the Background next reviews basic challenges confronting conventional, or what is sometimes referred to as analog, command and control performance and evaluation. Basic challenges considered include the "art" of command, the indirect relationship of command and control performance to mission outcomes, the nonlinear and context dependent value of information, as well as the complex and collective human-machine information chain that supports command and control.

In particular, two challenges associated with conventional command and control performance and evaluation are examined in some detail to substantiate the need for digital measurement methods. First, manually "burdened" measures and methods (e.g., recording, transcription, and correlation) rely too heavily on human observation, collection, reduction, analysis, and integration of command and control performance data. Second, analog media used to communicate command and control information, such as voice radio and paper maps, restrict information access and interaction that are essential to the conduct and evaluation of command and control. The relatively intractable nature of analog performance data contributes to human evaluation workload. The Background chapter concludes that burdened measures and analog media contribute to method inadequacies that range from biased reliance on subjective methods to failures in providing meaningful analysis, training feedback, and evaluation results.

The Findings chapter examines how the Army might apply digital technologies to evaluation with a focus on command and control performance. This chapter begins with a brief overview of how digital systems, especially Command, Control, Communications, Computer and Intelligence (C^4I) systems, are impacting training and evaluation in the Army. Attention then shifts to the focal issue of how digital systems impact command and control performance and particularly evaluation. This report's examination of digital command and control evaluation is divided into three parts or topic areas: Challenges, Opportunities, and Examples.

Challenges associated with digital command and control evaluation are complex and demanding. They include contemporary challenges of personnel turbulence as well as mission complexity and diversity. They also include many of the basic challenges associated with conventional systems and methods, as reviewed. However, the development and fielding of digital systems has resulted in a system-of-system complexity compounded by introduction and revision of numerous new and largely incompatible C^4I systems. Notably, the "productivity paradox" suggests that investments in technology require even greater investments in new organizational structures and processes, and these "other" investments are costly and take time. Moreover, the pervasive impact of digital information technologies on entire organizations and how they do business creates an expanding spiral of evaluation requirements.

Opportunities afforded by digital technologies, however, are unprecedented and can solve some evaluation challenges, including the ones they create. These opportunities are examined under five sections. First, a section titled Digital Media considers how the very nature of digital technologies represents new ways of doing business. Digital media capabilities are contrasted with analog media by their unparalleled ability to access information and act on information. Next, the Digital Instrumentation section examines how digital technologies are uniquely suited for automatically collecting performance data. The more C^4I systems become integral to performance, the more instrumented C^4I systems can record a log of all soldier-computer interactions correlated with the battlefield situation in which they occur.

Under the section titled Increased Scope and Precision, opportunities for collecting more exact and comprehensive measures of command and control performance are examined. For C^4I-based performance, digital technologies can collect data on any or all users at any or all times during a mission. Digital technologies can also readily sift voluminous databases for meaningful nuggets of information. More specifically, this section addresses how digital technologies can track and analyze psychomotor, procedural, cognitive, and collaborative performance. Under Meaningful Evaluation, the ability of digital technologies to situate performers and evaluators directly in the performance context is emphasized. The closing section on digital opportunities, Less Burdened Measurement Methods, examines how digital technologies might reduce the high workload associated with command and control evaluation.

Next, selected examples of digital measurement methods are provided and examined in the Findings chapter. These examples attempt to illustrate the potential of applying digital technologies to evaluation, and particularly to evaluations of command and control performance. Many of these examples are based on the Army's research and development efforts, with support from the U.S. Army Research Institute for the Behavioral and Social Sciences, and the remaining examples are drawn from related military research, including ground and air operational settings.

Despite the purported potential of digital measurement methods, the examples provided only suggest the ultimate potential of digital evaluation technology. Accordingly, many of the examples provided are currently just prototype concepts for command and control methods and measures that require further research and development. Overall, the examples provided are organized under three key issues that reflect the evaluative potential of digital technologies, and the requirement for additional research and development in order to realize that potential: data integration, data mining, and data visualization.

The Conclusions chapter stresses that the opportunities afforded by digital technology for improving training and evaluation are as yet only dimly envisioned (e.g., Bailey, 1996; Kurzweil, 1999). The need for improved performance and evaluation is great, however, and the opportunities must not be missed. Opportunity entails effort. The refrain "more research and development is required" applies in spades for digital measurement methods. Moreover, there are technical requirements and hurdles for improving performance and evaluation, particularly in the area of command and control. The Conclusions chapter identifies two *critical* technical requirements: instrumentation of all C^4I operational and training systems, and integration of digital technologies particularly C^4I systems and training simulations. The need to establish clear and compelling *standards* to ensure these requirements are met is also stressed.

This report underscores the need for a digital evaluation strategy, but it is not a blueprint or a handbook of digital measurement methods. That work remains, and requires a concerted and sustained effort by Army research, development, training, and evaluation personnel. As we learn by applying digital measurement methods and as digital technology advances, digital evaluation methods will evolve far beyond the limited examples examined here. Learning is in the doing, and as Binder (1964) reminds us, we must first "learn to sweat."

BACKGROUND

Measurement is integral to training and evaluation. This is true across the wide spectrum of research designed to establish the value of the object being evaluated. An object's value is relative to other similar objects, and to the viewer/evaluator of the object. Measurement provides a basis for relating objects and, ideally, a basis of interest to the viewing audience. An audience "internal" to the object, such as when a unit conducts an evaluation of its own training, might be interested in comparing its performance before and after training. An audience "external" to the object, such as when a unit or outside agency conducts an evaluation of some other unit's training, might be interested in comparing unit performance to normative or criterion standards. Overall, the purposes of measurement and evaluation, as used in this report, are manifold to include diagnosis, improvement, analysis, prediction, selection, and adjudication.

The intent of this report is to promote the development and application of digital measurement methods and automated measures of performance across the full spectrum of evaluation, particularly in the area of command and control. Digital measurement methods are a multi-purpose asset. The report's focus on applying digital technology to evaluation underscores the relatively unique measurement contributions afforded by digital technology including more precise and comprehensive measures of performance, more meaningful measures, and less burdened measurement methods.

From the outset, we stress that digital measurement methods are not a panacea, a cure-all, particularly in an area such as command and control. More traditional measurement methods, such as observation, interviews, surveys and questionnaires are still needed. However, many perceived shortcomings in the use of more traditional measurement methods can be offset by digital measurement methods. Generally, more balanced measurement methods are needed that include both traditional and digital measurement approaches. Specifically, this report focuses on the use of digital measurement methods to improve command and control performance and evaluation, and to correct an over reliance on subjective measures about performance.

Army Evaluation Requirement

This section begins with a brief review of military and academic training and evaluation literature that discloses a pervasive requirement for more efficient and effective evaluation methods. The need for improved evaluation methods is not unique to the Army. Members of the behavioral science community, particularly trainers and researchers in the area of computer-mediated performance, as well as educators and business professionals, currently face the same challenge.

The Army's tradition of exacting performance standards, proficiency certification, and performance-based training and evaluation is receiving new emphasis (Brown, 1999; Rosenberger, 1999). The Army's foundation training document, Field Manual 25-100, asserts that evaluation is integral to training (U.S. Department of the Army, 1988). This capstone document stresses that the purpose of training evaluation is to measure the demonstrated ability of individuals, leaders, and units against specified training standards. Accordingly, the Army's training literature uniformly reinforces this evaluation requirement, as indicated in U.S. Army Training and Doctrine Command (TRADOC) Regulation 350-70 (U.S. Department of the Army, 1999). This regulation directs training developers to rigorously follow the Army's evaluative methods and procedures. Core evaluation methods used by the Army are described and referenced in this guide. These methods include the Training and Evaluation Outlines recently made available in the Automated Systems Approach to Training (ASAT), as well as the task, condition, and standard formats exemplified in the extensive Army Training and Evaluation Program (ARTEP).

Despite this tradition and a highly codified methodology, contemporary factors may be undermining the Army's foundational methods for conducting training and evaluation. The most telling indication of cracks in this foundation comes from the observations of Colonel John Rosenberger at the National Training Center (NTC), the Army's most demanding warfighting arena (Rosenberger, 1999). In essence, the combined arms teams that rotate through the NTC evince a trend of declining readiness relative to the center's standing opposing force (OPFOR). Rosenberger highlights two factors, in particular, contributing to the problem: annual turnover rates greater than 40 percent across units, and more complex warfighting.

Brown's assessment of the problems contributing to the Army's current concerns about training and evaluation, sustains the conclusions of Rosenberger (Brown, 1999). While personnel turnover complicates the development of habitual team organization and performance, the geographic dispersion of personnel does also. For example, Brown estimates that a typical Army National Guard Separate Brigade rotates to the NTC with units from over 20 states. Brown also acknowledges the increased complexity of warfighting and contends this complexity is compounded by the increased diversity of the contemporary mission environment. Diversity factors include significant expansion in the number and types of missions the Army must prepare for, more complicated joint and combined organizations, and an increasing mix of conventional and digital systems.

Evaluation shortcomings may be the most influential and discriminating reason why the OPFOR units at the NTC increasingly outperform the incoming rotational units, Rosenberger concludes. "Unlike the units they face, the OPFOR *confirms* that every solider and every leader possesses the knowledge, skill and ability to perform his/her duties *before* they are permitted to fight..." (Rosenberger, 1999, p. 9). Notably, multiple methods of evaluation are used by the OPFOR to include: a series of written exams, oral exams, terrain walks, apprenticeships, and hands-on demonstrations.

Inadequate evaluation and its detrimental effects on NTC performance may be most pronounced at the leader levels, particularly the command and staff personnel central to command and control.

> *...combined arms battalion and brigade commanders are not required to prove or demonstrate a mastery of battle command skills and tactical competence before being placed in command. It is not, and has not been, a prerequisite for command selection. It shows at the NTC, year after year.* (Rosenberger, 1999, p. 10)

Rosenberger argues that the Army must restore or create the conditions to achieve full combat potential, and that the key condition is the opportunity to train at the frequency required to meet and sustain performance standards. Achieving this condition amidst counter conditions such as personnel turnover and warfighting complexity requires more effective and efficient training and evaluation strategies and methods, Rosenberger contends.

Brown's work on the formulation of Army training, and particularly digital training, also underscores the Army's evaluation requirement of insisting on proficiency. His enduring advocacy of evaluation stresses the fundamental precept of the Army that all learning is evaluated, and all evaluation is learning. This precept drove the Army's formulation of performance-oriented training based on tasks, conditions, and standards. This formulation became the building block for cornerstone programs such as ARTEP and Skill Qualification Tests (Brown, 1999). Brown asserts that most of this evaluation emphasis is gone today, and that current evaluation is generally self-generated or decentralized to the personal preferences of each small unit commander.

Brown's analysis underscores contemporary inadequacies in the evaluation of command and control. His assessment concludes that little evaluation content is available for certification of leader or staff team proficiency in the exercise of control, and even less is available on the exercise of tactical command. In summary, he states: "Reinstitution of evaluation of tactical proficiency...is arguably the single most important action to be taken to improve the quality of tactical leaders today" (Brown, 1999, p. 11).

A call to standardize evaluation methods and performance measures at the Army's Combat Training Centers (CTCs) was recently issued (General Accounting Office (GAO), 1999). That report concluded that neither the Army nor individual units are achieving the full benefits of CTC training for reasons including: units are ill prepared upon arrival, unrealistic training, aging equipment, and an inability to capitalize on lessons learned from the centers' exercises. Specifically, that report's conclusion on lessons learned stressed the Army has not fully implemented a plan to adequately assess training, including training in the area of command and control.

The GAO noted that because no standardized data collection programs exist at the CTCs, the Army could not develop lessons learned from the exercises. The GAO, therefore, urged the Army to implement a comprehensive plan for collecting, analyzing, archiving, and disseminating CTC data. The report noted that such a plan was developed in 1995 by the Combined Arms Center, but not funded. The GAO also stressed a fundamental evaluation weakness: the Army applies no standard performance measurement set to provide benchmarks that might better focus training and gauge readiness. In conclusion, the GAO report noted that performance measures are fundamental to management and consistent with the Government Performance and Results Act that requires that all government programs be evaluated with performance measures (Office of Management and Budget, 1993).

Behavioral Science Evaluation Requirement

To address the military requirement for better evaluation methods, a brief review of evaluation literature from the behavioral sciences was conducted, particularly in areas of training evaluation and computer-mediated work. Based on this review, the authors of this report conclude that the Army's need for better evaluation methods is in many ways shared by the behavioral science and education communities, and much work remains. For example, while the National Research Council noted some promising evaluation strategies and methods in a recent review of training evaluation, their overall assessment of this area disclosed more problems than answers (Nickerson, 1995).

Three basic and unresolved training evaluation issues identified in the Council's review were that evaluations were far too often piecemeal, irrelevant, and subjective. "A central methodological problem is the lack of a comprehensive array of performance measures that are acceptable to both researchers and practitioners" (Shields, Cavallaro, Huey & Van Cott, 1995, p. 92). A more comprehensive approach is needed that assesses the multidimensional aspects of learning, including behavior, cognition, and affect. However, the variety and specificity of job tasks in training research and the lack of criteria for doing comprehensive program evaluations results in disparate and fragmented measures and methods. To avoid a perceived pattern of measurement irrelevance, the Council urged that more logical and convincing evaluation methods be tied to formal assessments of actual worker needs and job requirements. Moreover, the Council's review confirmed a reliance on subjective measures, particularly crude ratings and informal interviews, in most work settings.

The need for more objective and balanced measurement methods is underscored by persistent shortcomings in research directed at cognitive and collaborative performance (e.g., Ehret, Gray & Kirschenbaum, 2000; Salas, Rhodenizer & Bowers, 2000). A recent review of the research on aircrew resource management training, for example, concluded that research had not impacted actual training to an acceptable degree (Salas et al., 2000). The recommendations they provided for ensuring greater impact stressed the requirement that researchers apply more advanced technology to enhance the value of collaborative, crew-level training. The key technology requirements identified in their recommendations were digital measurement methods, such as eye tracking and speech recognition, coupled with simulation-based environments. Salas et al. stress such technologies are needed to measure team processes, and to provide more specific and immediate performance feedback.

Similarly, a review of computer-mediated work, closely related to digital command and control performance, disclosed pressing and unresolved evaluation issues. Notable issues include the increased cognitive content of work and the emphasis on group work. As the intellectual basis of work expands, the National Research Council detected growing inadequacy in the traditional task analytic methods devised to specify the displays and controls needed for operators (Nickerson, 1995). Moreover, methods are sorely lacking for evaluating complex cognitive and social skills. The Council stressed that methods are needed to understand how professionals perform their cognitive work and why they perform it in that manner. For example, despite the demonstrated improvement provided by some automated decision aid systems in the medical profession, they are not used. The Council concluded that the evaluative need is not so much for summative assessments on the merits of job aids, but for formative methods to improve aids and ensure they are used.

The collaborative nature of computer-mediated work complicates evaluation, particularly in what the Council refers to as fluid, versus restricted, job domains. In fluid domains, such as knowledge-based work, the boundaries and organizational structure are less concretely defined. Even when defined, they must accommodate the unpredictable growth of information. Methods for evaluating such work domains and the task requirements to operate in them are much needed: "...no one has yet devised a method that is easily learned and acceptable to practitioners as well as researchers" (Nickerson, 1995, p. 95).

In summary, our review for this report of evaluative methods from behavioral science literature applicable to the Army's emerging evaluation requirements was not very productive. While the evaluation challenges are similar, particularly for computer-mediated work in complex and collaborative fluid domains, evaluation methods are wanting. We concur with this National Research Council assessment: "A major practical need is for significant improvement in the means for performance assessment" (Nickerson, 1995, p. 25).

At best, we conclude that better evaluative methods are emerging, and that their emergence is linked closely to advances in digital technology. For example, despite the Council's dismal conclusions about current evaluation methods, they provide two notable recommendations for improving evaluation through the application of digital technologies. First, the Council recommended consideration of evaluation methods that view cognitive functions as similar to computer programs. As computers cannot function without a protocol or program, similarly a protocol is needed to describe the overt and more covert cognitive type tasks that make up a particular job. The Findings chapter examines the opportunities provided by digital technologies to track and analyze behavioral protocols for psychomotor, procedural, cognitive, and collaborative performance.

Second, the Council's summary report acknowledges "...the unprecedented capability of computer-based systems to represent information in graphical or pictorial format" (Nickerson, 1995, p. 41). The Council's discussion of this capability focuses on the need to identify the tasks which might best use visualization tools and find better ways to evaluate how well visualization-aiding systems help users perform their tasks. The Findings chapter concludes with an examination of how digital technologies can and might provide data visualizations that support command and control evaluation.

Conventional Command and Control Evaluation

> *The literature which deals with command and control performance research and measurement methodologies is sparse, scattered, and of highly variable quality.* (Crumley, 1989, p. vii)

More than a decade after Crumley's assessment, we conclude that the status of research on command and control performance, particularly measurement methods, is not much improved. Why? Primarily, because command and control evaluation is beset with challenges. Some of the key challenges associated with conventional command and control evaluation are examined below. They focus particularly on the use of analog command and control equipment and procedures, equated in this report with the term "conventional." And we stress that a more thorough discussion of such evaluation challenges is ably documented in other efforts (e.g., Alberts, Garstka & Stein, 2000; Allen, 1999; Crumley, 1989; Halpin, 1992).

More basic challenges examined here include the very nature and definition of command and control, and its indirect effects on mission outcomes. Additional evaluation challenges include the context-specific value of information and decisions, and the complicated collective chain of human and technical processes that support command and control. The following review focuses most closely on two issues that severely challenge these evaluations: the high workload associated with manually "burdened" methods and measures, and the limitations imposed by analog media.

Challenges

The evaluation challenge begins with understanding the purpose and role of command and control. While formal definitions vary, a succinct and relatively clear definition is that "command is the authoritative act of making decisions and ordering action; control is the act of monitoring and influencing this action" (U. S. Department of the Army, 1997). A definition more directed at evaluation purposes is that "command and control is the management component of any system (Finley, Muckler, Gainer & Obermayer, 1975). Adequate evaluation requires consideration of the command and control *sub*system in relation to the overall military system. As a convenience, this report uses the term "command and control system" (versus *sub*system), and this usage includes descriptions and quotations from the literature reviewed. Despite this convenience, the relation of command and control to the overall system is reinforced here and throughout the remainder of this report.

The functions of command and control are performed through an arrangement of personnel, equipment, communications, and procedures employed by a commander in planning, directing, coordinating, and controlling forces and operations to accomplish a mission. The personnel, equipment, communications, and procedures that compose a command and control system must provide decision specific information to the commander. As system manager, the subsystem of command and control establishes data acquisition needs and data processing procedures for the entire system that meet the commander's information requirements.

Challenges in evaluating command are underscored by doctrinal and related literature that stress command is an "art." The art of command includes assigning missions, prioritizing resources, guiding and directing subordinates, and focusing the entire unit's energy to accomplish clear mission objectives (Bornman, 1993). A notable instance of this art is battlefield visualization, the process whereby the commander understands the current state of the battlefield, the desired end state, and the sequence of activities required to achieve the end state.

In contrast, the "science" of control is considered more amenable to evaluation, more measurable. The science of control entails defining limits, computing requirements, allocating resources, prescribing report requirements, monitoring performance, identifying and correcting deviations from guidance, and directing subordinate actions to accomplish the commander's intent (Bornman, 1993). For comparison, envisioning a course of action (COA), the sequence of activities required to achieve the end state, represents the art of command. Assessing a unit's performance relative to a prescribed COA including its end state represents the science of control. As Bornman notes, control adjusts unit actions to match a COA; command adjusts the COA to achieve the end state.

The visionary aspect of command as art suggests that command must be subjectively evaluated by highly qualified and skilled instructors and mentors (Brown, 1999). Reinforcing the elusive nature of command performance is the reality that no two actual battlefield situations are identical. No matter how many warfighting experiences or command training exercises a commander has experienced, the current situation is always unique and optimal performance unknown (Allen, 1999). Allen also stresses that a commander's decisions, a mainstay measure of command performance, are only indirectly related to the end states achieved. The efficacy of command, as measured by decision making, is heavily influenced by control factors, including the commander's ability to adjust operations, as required, to meet the desired end state or mission outcome.

Seminal work on the evaluation of command and control performance was conducted in the 1970's (Olmstead, Christensen, & Lackey 1973; Olmstead, Elder, & Forsyth, 1978). This work examined and clarified the indirect relationship between command and control performance and mission outcomes. As reviewed by Crumley (1989), the 1973 work by Olmstead and colleagues demonstrated that command and control performance accounted for over 86 percent of the variance in command *decisions*. In contrast, their 1978 work demonstrated that command and control performance accounted for only 40 percent of the variance in *outcomes*, or end states achieved.

Two factors, in particular, were cited by Crumley to explain these results. First, many intervening variables complicate the relationship between decisions and outcomes, especially in battle. Second, the latter work relied primarily on subjective measures, namely observer ARTEP ratings of command and control task performance. In contrast, the robust relationship found in the earlier work relied on a detailed and objectively measured analysis of the command and control process and product, without regard to battle outcomes. The process measures included all written and oral communications among the command and staff personnel, and the product measures were the decisions made by commander and his staff.

Another challenge in evaluating command and control performance is how to determine the value of information. Information is the basis for both command decisions and control over how those decisions are executed and adjusted. However, there is no generally accepted methodology for evaluating information (Reidel & Fitz, 1986). The value of information is nonlinear and context-dependent. This is especially true for the context-specific conditions unique to any battlefield situation or state. In a more balanced situation, Allen (1999) suggests, any additional piece of information might tip the scale for an opposing force. On the other hand, a force that achieves actual information dominance might imbalance the situation so much that their opponents never really benefit from any additional information they might obtain.

Numerous factors complicate efforts to establish the value of information. A more comprehensive assessment of information value, for example, might include factors such as information timelines, accuracy and relevance, the volume and cost of information, and ultimately its use (Reidel & Fitz, 1986). The value of any particular piece of information is dependent upon preceding and subsequent information, and upon the performance of all command and control elements before and after the information was received. Moreover, commanders are trained to make decisions in the absence of information, and may make the correct decision even without key information, or the wrong decision with the right information (Allen, 1999).

Another major evaluation challenge is the complicated collective chain of human and machine processes that support command and control. Conventional evaluation approaches attempt to untangle this chain by identifying and measuring human versus machine components and processes separately. The work of Metersky (1986), for example, proposed an evaluation strategy in which separate human and machine measures of performance are considered, including separate machine measures for command and control equipment versus weapon systems. However, evaluators often contend it is nearly impossible to measure the benefit of a single command and control asset due to the connectivity and interdependence of the overall command and control process (Allen, 1999).

A final set of challenges considered below are the barriers and burdens associated with observing, collecting, and assimilating command and control performance data. While direct observation and objective measurement of performance during military training exercises are always difficult, they are particularly so for command and control. The essence of command entails internal information processing activities, such as risk assessment and decision making, that are usually covert and unavailable to observers. Moreover, the perceptual and cognitive activities that underlie command, such as pattern recognition and battlefield visualization, are often automatic processes "unavailable" to performers.

In contrast, the external processing of information, including communication, which underlies command and control performance is generally observable and recordable. However, measurement and particularly evaluation of this external information is beset with challenges. Two of these challenges that underscore the need for digital measurement methods are examined in the following sections: burdened measurement methods and analog media limitations.

Burdened Measurement Methods

The actual measurement of conventional command and control performance is a task invariably performed by human observers. Job aids and tools, such as observation checklists and audio and video recordings, support the observers' measurement efforts. Despite this support, however, measurement methods routinely used for evaluating conventional command and control performance and training are manually "burdened" measures. By far the majority of the Army's assessments of command and control performance are made to support training rather than test and evaluation efforts. For training assessment, the measurement burden begins with the multiple roles and responsibilities assigned to observer personnel who are generally referred to as observers/controllers (OCs).

The multiple tasks required by OCs, in addition to observation and data collection, are only suggested by their dual hat title. Perhaps, the most thorough and current analysis of OC task responsibilities is that provided by Brown, Nordyke, Gerlock, Begley, and Meliza (1998) in a study of training analysis and feedback aids to support live training. Their analysis describes an OC as a tactically and technically competent officer or noncommissioned officer who performs multiple roles including trainer, observer, and exercise controller. Brown et al. also identify the numerous, diverse, and conflicting tasks inherent to the roles of OCs. The heavy OC workload associated with data collection and evaluation is noted in this study conclusion:

"...the analysis clearly shows that OCs are involved in data collection tasks which divert them from player behavioral observations, coaching and mentoring" (Brown et al., 1998, p. 52).

In short, OC roles and responsibilities are often so extensive and intensive that the actual collection of data can be considered a diversion. Admittedly, OCs at the Combat Training Centers are provided data collection support from partially instrumented ranges and analytic support from personnel in Training Analysis Facilities (TAF). However, these data collection assets are diminished, and often nonexistent, for training at home-stations and institutional settings. Moreover, regardless of setting or assets, OCs are the Army's principal training evaluators and front-line data collectors.

To better appreciate why manual measurement methods are a burden, we next examine two basic issues. Why OCs collect data? What types of data OCs collect? The main reason OCs collect data is to provide training feedback to the immediate training audience, such as a rotational unit at the maneuver CTCs. Another reason for OC data collection is to assess and document unit performance for external audiences, such as TRADOC's Center for Army Lessons Learned (CALL). The reason for data collection, of course, influences the types of data collected. First, we examine the OC data collection and analysis requirements for the training audience, based on the analysis by Brown et al. (1998).

For the training audience, OCs collect data to provide intrinsic and extrinsic feedback. Intrinsic feedback is provided to the training audience during the exercise. This feedback represents the sensory stimuli that should be inherent to battlefield conditions and soldier activities, such as flash, bang, and damage effects from weapon firing. While tactical engagement simulation provides some of this stimulation, OCs must provide much of this feedback in their role as exercise controllers to increase training realism. The OC data collection requirements for providing intrinsic feedback are substantial. To simulate the effects of a single minefield explosion, for example, OCs need to collect data on minefield location, type, density, width, length, and orientation.

The extrinsic feedback provided to the training audience by OCs includes coaching, Take Home Packages (THPs), and After Action Reviews (AARs). Data collection for coaching relies mainly on informal observations by the OCs to provide feedback during all phases of an exercise, including planning, preparation, and breaks in the execution phase. Data collection for THPs and AARs includes a wide range of data types to provide feedback after a training exercise. The types of data required include categories such as: network, radio communications, observation, planning products and terrain information. The analysis by Brown et al. (1998) discloses that the majority of this data, except for electronically collected network data, must be collected by OCs and supporting analysts.

The OCs not only collect much of the data for extrinsic training feedback, they play the primary role in analyzing this data and preparing the feedback materials required for THPs and AARs. This training analysis requires a mastery of tactical and technical issues best ingrained from past training and warfighting experiences. It also requires detailed knowledge of and reference to a substantial library of doctrinal and training literature. In sum, even after completion of a training exercise the OCs primary job of providing training feedback to a unit has just begun: “Preparing detailed observations, translating observations into key issues, and linking key issues to exercise objectives and military doctrine are intensive and highly time-consuming tasks “ (Brown et al., 1998, p. 56).

Second, OCs are the primary data collectors for training evaluations conducted for external audiences. Their prominent role in such evaluations is underscored by a recent RAND report titled *Determining Training Proficiency at Combat Training Centers: Data Collection Instruments* (Hallmark, Crowley, Leonard, Lippiatt & Sollinger, 1998). These data collection instruments consist entirely of questionnaires completed exclusively by OCs at specific times during a unit’s rotation at a CTC. Each of the 29 questionnaires included in this instrument set is customized to a specific organization in a brigade team. Each questionnaire addresses five functional areas such as maneuver and artillery, and the command and control performance of every combat organization. The design of these questionnaires reflects evaluator efforts to reduce the observation, and particularly data collection, burden on OCs: “Considerable effort went into minimizing the burden of filling out the questionnaires” (Hallmark et al., 1998, p. 6).

The RAND report provides some background on their questionnaire design that amplifies this section’s emphasis on OC workload and its potential impact on observation and data collection. The evaluation instruments in traditional ARTEP Mission Training Plans (MTPs) typically include for each organization 30 to 60 tasks, with 10 or more subtasks for each task. For a battalion task force’s 53 MTP tasks, the data collection checklist requires 163 pages, and for a company team’s 47 MTP tasks, the data collection checklist requires 330 pages. The approximately 55 evaluation items included in each RAND questionnaire, therefore, were based on a very selected set of critical subtasks from these MTP instruments.

As indicated, questionnaire design stressed a minimal burden on the already busy OCs. To ease OC data recording, each item was linked to a 5-point performance rating scale in which 1 = “Not sufficient” to 5 = “Superior” (0 = “Not done, but should have been”). Reportedly, items were limited to those behaviors directly observable by an OC. Another practical constraint was to make each questionnaire short enough to be printed on an OC “card” sized to fit the pocket of a battle dress uniform. According to the RAND report, OCs began completing these questionnaires at selected CTCs in June 1998. Despite the laudable effort by RAND, such external audience requirements only increase the OCs evaluation workload.

Even for more formal evaluations, such as testing and evaluation efforts, the observation and measurement burden still falls heavily on humans. This is especially true for evaluations of command and control performance due in part to the connectivity and interdependence of the overall command and control process. Unlike the multiple roles played by OCs for training evaluations, however, observers for more formal test and evaluation efforts are often solely dedicated to the task of observation and data collection. The potential reduction in workload for dedicated observers and evaluators, however, is invariably offset by increased requirements for more detailed and comprehensive data.

A good example of how evaluation workload tends to increase is the development of the Army Command and Control Evaluation System (ACCES) (Halpin, 1992). The ACCES measurement system was designed to evaluate the effectiveness of command and control as a function of the headquarters staff rather than force effectiveness, or battle outcomes. Reflective of the difficulty of evaluating command and control, nearly a decade of work was spent on development and refinement of the ACCES methodology. During this development, ACCES repeatedly grew in order to collect and report "...more and more detailed information..." (Halpin, 1992, p. 25). To avoid getting lost in such detail, new "primary" measures were later identified and added to ACCES, which were mainly aggregates of prior measures. However, this refinement of ACCES raised the observation and data collection burden.

Measurement design in ACCES stressed objective measures of the process and outcome of command and control performance, rather than subjective estimates about the "quality" of a product or decision. The most essential and difficult ACCES data collection was human tracking of information exchanges. To prepare for data collection, and subsequent data reduction and analysis, ACCES included a Program of Instruction (POI) for data collectors. This POI included 14 lessons that began with overviews of ACCES and Army Command and Control, progressed to exercises for recognizing key events and documents to be recorded or collected, and closed with data reduction procedures. Halpin (1992) describes the difficulties and costs associated with maintaining a cadre of experienced data collectors for ACCES. This description applies to almost any command and control assessment team, including the OCs at the CTCs.

A short examination of ACCES data collection procedures may convey the observation and collection burden. During each headquarter exercise, approximately 20 ACCES data collectors were stationed at "critical" command and control nodes, provided with clipboards and data sheets, and asked to record all relevant information exchanges. Based on their prior military expertise and the ACCES training, the data collectors were required to recognize exchanges and events related to the data of interest and record this information in their chronological journals. Periodically they were required to also record key information from posted map overlays and status boards at their node location. Special data collectors stationed at the simulation center for the exercise recorded all critical tactical and training events throughout the exercise to maintain the overall context of the mission. This context included capturing simulation "ground truth" for later assessment of the unit's timeliness and accuracy of information exchange, and the unit's accuracy in forecasting future battlefield situations.

After the exercise, the arduous ACCES data collation processes integrated the data from each collector and the "ground truth" context. An example from this process may portray the challenging task of relating disparate data points collected across command and control elements into a meaningful pattern or finding. To assess the decisions to withdraw a force because help will not arrive in time and commit artillery to secure more time, Halpin notes the following process:

> *During data reduction and collation, these decisions are noted, and that observer's notes, and any others which are relevant, are scanned to determine the circumstances of the decisions. What information did the decider [sic] have available to him? Was it accurate, timely, complete? What interpretations or understandings had been stated? What predictions had been made? The data are also scanned to the consequences of the decision. How long did it take before an order was issued? Did the order match the decisions? Was the order unambiguous? Or did subordinate commander request*

clarification? If clarification was requested, how long a delay ensued before the clarification was provided? Were the subordinate commands allowed enough time between issuance of the order and the scheduled execution time for them to do their own planning and preparation?" (Halpin, 1992, pp. 20-21)

Drawing valid conclusions about performance requires a full consideration of the context in which it occurs. The connectivity and interdependence of all command and control elements and components results in an extended behavioral context. Many command and control components, including the staff, are simultaneously exchanging and processing information in a shared context. As Halpin points out, in some cases different plans and decisions are being made at the same time in different locations or contexts.

In summary, human observation, collection and collation of command and control data is challenging and time consuming. In current CTC and home station training environments, the observation and data collection burden diverts attention from more critical aspects of training, such as coaching, mentoring, and providing feedback. Unfortunately, feedback requirements for external audiences invariably increase this burden. By necessity, more formal and robust evaluations of command and control performance are usually supported by dedicated data collectors and analysts.

As noted, however, potential reductions in workload are offset by the need for more detailed and comprehensive data and meaningful analysis. As a consequence, the time required to manually integrate and interpret more detailed and objective data from disparate sources adds substantially to the cost and, perhaps, the sparse and uneven quality of command and control performance evaluations. A related consequence is that lags induced by manual integration and interpretation of more detailed and objective data on command and control performance prevents timely and meaningful feedback to the primary training audience.

Analog Media Limitations

A final but focal challenge considered here is the limitation imposed by the analog media on the performance and evaluation of command and control in a conventional environment. All media are essentially communication channels. They are technologies, or tools, used to store, process, and communicate information. This section examines how the passive and disjunctive nature of analog media restricts the interaction and collaboration essential to both the conduct and evaluation of command and control. Analog media also increase the mental and physical workload of command and control performers and evaluators. First, a description of analog media characteristics in relation to command and control is provided. Then, the impact of analog media on command and control performance and evaluation are examined. As preface, we recall McLuhan's assessment about the fundamental and pervasive impact of media: "... the medium is the message because it is the medium that shapes and controls the scale and form of human association and action" (McLuhan, 1964, p. 9).

The analog media used by the military to communicate command and control information are diverse and numerous. Even a partial list of traditional analog media includes: horns, drums, whistles, fire, smoke, animals, flags, rocks, sand tables, chemical lights, binoculars, hand and arm signals, maps, acetate overlays, grease pencils, map boards, status charts, typewriters, report forms, journals, radios, teletype, telephones and cell phones. Many of the Army's analog media are being converted or transformed into digital communicators. However, a predominant segment of the Army is, and for some time will remain, a conventional force reliant on analog media. In fact, analog refinements are ongoing and the nature of these useful, but modest, refinements only underscores analog media limitations. More recent analog refinements include the use of "Post It" notes for annotating Spot Reports to maps (Herl, 1999), and more innovative flag signal markings for casualty evacuation (Nolan, 1999).

Analog media are passive "carriers" of data. While analog media can collect, store, reproduce, and transmit data; they often do not, or do so in an imprecise manner (Gates, 1996). For example, most analog media cannot error check and correct the data they transmit. They are not what are commonly referred to as "self-healing" media that can remove extraneous static, hiss, and noise or recollect and reassemble missing data. The data value conveyed by analog media is approximate, based on analogy. The more often analog data are transmitted, the less accurate it becomes. Moreover, analog media do not generally act on or process the data they carry. Usually, humans must "act on" analog data to transform it into information. The informational aspects of analog data must generally be encoded by human senders and decoded by human receivers. An iterative process of transforming data into information, and information back into data, is a burden imposed by the passive nature of analog media.

The impact of the passive nature of analog media on command and control performance is severe. The medium of voice radio, for example, requires users to repeatedly transmit information received over one combat radio net to other nets. Often users must store or record transient voice data on note pads or map sheets and then transmit the information to other users who, in turn, repeat the store and transmit functions. Moreover, analog media such as radio do not act on the information to support or augment it. For example, voice radio requires users to perform identification and authentication procedures which account for over one-half of the "information" transmitted during training at the NTC (Phelps & Kupets, 1984).

Analog media are also stand-alone or disjunctive information carriers. They do not readily exchange information between media types or with other analog systems. For example, the receipt of an analog voice radio message providing the location of an enemy target cannot automatically slew and fire a weapon system's gun at that target. They are not what are commonly referred to as "seamless" media joined by a common, uniform standard that regulates their control and translation (Kittler, 2000). Rather, analog media signals and formats are often individually standardized and incompatible with other media, that is they are medium specific.

Limitations imposed by the disjunctive nature of analog media on command and control performance are numerous and costly. The exchange of information between voice radio and paper map sheets, for example, requires human intervention. Most command and control communications must convey spatial data about the battlefield. Voice-based systems require that soldiers repeatedly encode spatial data, such as map coordinates, into alphanumeric formats called grids at the sender station and then decode these same elements back to spatial formats at the receiver station. More complex communications of map overlays, acetate blueprints of the operation, are exceedingly difficult to transmit by voice. The "transmission" of detailed overlay information, therefore, is rarely completed over voice radio and generally requires hand copying and face-to-face physical distribution.

The passive and stand-alone nature of analog media formats often results in information that is static and decontextualized. The military's analog media formats rely heavily on alphanumeric characters and tabular formats (Gerhardt-Powals, 1996). For example, a table format called an execution matrix is used to coordinate unit activities. A simple example of an execution matrix might be a 5 x 5 table that prespecifies for each of five units, their expected locations and activities for five successive mission segments. The inherent stability of an analog table format (Kozma, 1991), however, may not readily convey the dynamic patterns implicit in an execution matrix. Although text and table formats afford precise and detailed information, they are often decontextualized from relevant aspects of the task environment, such as battlefield space and geometry for command and control tasks.

Overall, the passive and disjunctive nature of analog media substantially limits the information processing and communicating functions essential to command and control performance. The limitations of analog media not only increase command and control workload; they also dictate to a large degree what work is required and how it must be performed. Such effects are strikingly evident in an examination of the analog transformation of command and control for artillery.

Per-Arne Persson (1999) notes that with the introduction of long-range artillery in the later half of 19^{th} century, revised fire control procedures were sorely needed for targeting and effects assessment beyond line-of-sight. Artillery devised an exacting set of analog fire control procedures from the forward observer back to the artillery batteries to ensure safety and reliability of effects. Changes to such procedures, due to attrition for example, were most difficult: "...reallocation of permissions and rights to fire was a tedious affair" (Per-Arne Persson, 1999, p. 56). Later, the introduction of digital technologies and media forced a reassessment but only a partial revision of these controls.

The impact of technology on command and control is profound and pervasive: "...the meaning and content of what is 'work' has been standardized, not because it is the best way, but instead according to what standard technology proposes and admits" (Per-Arne Persson, 1999, p. 56). Analog media not only limit command and control performance, they also instill a pattern of internal control, of work, that organizations such as the military are slow to abandon despite changing conditions.

Analog media also pose a severe challenge to command and control evaluation. The limitations of analog media increase the workload associated with nearly every phase of evaluating command and control training and performance. Recall, the OC workload analysis by Brown et al. (1998) examined the severe impact of analog media on command and control evaluation. To summarize this impact, Table 1 provides sample descriptions of how key evaluation tasks are performed at the CTCs, based on edited excerpts from Brown et al. (1998). As indicated in Table 1, OCs and supporting analysts can eavesdrop on and electronically record voice radio communications related to command and control during CTC exercises. However, the actual collection and collation of this voice data is time consuming and labor intensive. The passive and transient nature of voice data requires that OCs and analysts, like other receivers, must often transcribe voice messages on note pads, observation sheets or maps during the exercise. And adequate collection of this voice data requires recording both message content and related data on sender, receiver(s), and time. The task description for data recording and reporting in Table 1 indicates how analog media increase task workload. Tracking the flow of voice data across the unit is a completely manual task that requires access to multiple tactical nets. Tracking and integrating a series of related voice communications, even relays of the same message, is even more complicated. Such tracking requires extended "cross talk" between OCs and analysts who each monitor different tactical nets.

In addition, the disjunctive nature of voice radio requires OCs and analysts to repeatedly transform and plot alphanumeric location data onto map and overlay media. A more complex task, such as battle tracking that entails the continuous and collective flow of voice radio information, requires OCs and analysts to repeatedly perform manual observation, collection, recording, and integration tasks.

Analog paper and acetate products such as maps, orders, and overlays also contribute to evaluation workload. As indicated in Table 1 under Data Collection and Observation, OCs must manually collect orders and overlays periodically from numerous locations and then hand carry them to supporting analysts. Almost any analysis of the information conveyed by analog overlays and orders is difficult and tedious. For example, checks on the internal consistency of these products requires evaluators to manually compare text and graphic content on the multiple hand-made copies prepared by numerous exercise participants during the exercise. And analyses directed at relating information from different mediums, such as overlays and voice radio reports, requires OCs and analysts to manually collate, integrate, and compare this information.

Analog media also complicate evaluation efforts to relate command and control data to actual exercise conditions. Assessing the accuracy, completeness, or timeliness of the information conveyed by voice radio or overlays and orders, requires manual comparison to tactical conditions and activities during the exercise. Even though range instrumentation at the CTCs automatically provides data on both friendly and enemy vehicle locations and actions, this data must be manually compared with the analog voice and paper command and control data. Finally, as indicated in Table 1, the analog nature of the OCs large library of training references further complicates efforts to assess performance and provide feedback.

For more formal evaluations of command and control performance, analog media limitations contribute heavily to the evaluation workload and restrict the precision and scope of the data collected. By design, more formal evaluations should result in a more robust and comprehensive database than that normally expected from training evaluations. Such expectations require a commensurate increase in evaluation workload, particularly given the analog media used for conventional command and control. The basic and manually intensive tasks of data collection, analysis and integration are the same as for training evaluations, but many additional subtasks are added and standards are more stringent.

Table 1

Examples of Evaluation Workload Associated with Analog Media Limitations*

Evaluation Task	Workload Description
Data Collection and Observation	Observer Controllers (OCs) and Training Analysis Facility (TAF) analysts at the Maneuver CTCs have the capability to eavesdrop on voice communications during the exercise. The TAF has the capability to record, time stamp, and playback voice communications on all BLUFOR tactical nets to support AAR presentations. (p. 46) OCs collect orders and overlays from BLUFOR and ensure the timely delivery of these items to the TAF. The OCs also record and report significant tactical events, battle damage assessment and information pertinent to effectiveness of BLUFOR's reconstitution efforts. (p. 51)* Observations of human behavior and coaching may require the OC to separate himself from his vehicle and controller communications. This presents a dilemma for the OC. If the OC is unable to hear others calling him on the radio, his absence from the control net may impact adversely on the control of a critical exercise activity. If he stays with his vehicle to respond to radio calls, he may miss an observation or coaching opportunity; i.e., generation of courses of action by the Bn TF staff within the BLUFOR tactical operations center (TOC). (p. 53)
Data Recording and Reporting	There are numerous recurring, pre-formatted reports that OCs submit to their senior OC or a TAF analyst.... After finding the desired report in the appropriate standard operating procedure (SOP) or in a compilation of reports collected from several SOPs, the OC hand-writes the report and/or submits the report orally over a control net. The receiver of the report (another OC or TAF analyst) records the information by hand. On occasions, the nature of the report may require the OC to travel to the addressee's location to deliver the report. In any case, preparation and submission of reports is a manually-intensive activity..... (p. 55)
Data Analysis and Integration	At the end of the exercise, OCs review their observations, group their observations into key issues for discussion, then link the key issues to doctrine and battle outcome. (p. 51) The TAF analyst reviews OPORDs, overlays, FRAGOs, requests, reports, free text messages, and warnings.... (p. 48) OCs must transport a large library of references to perform the tasks outlined above. Ensuring all OCs have a <u>current</u> reference library requires a considerable effort by the OC team's senior leadership. Manual searches for information in paper-based references is terribly time-consuming and often results in an incomplete search. (p. 54)

*Note. Examples based on edited excerpts from the training aids analysis conducted by Brown, Nordyke, Gerlock, Begley, and Meliza (1998) with page numbers in parentheses. BLUFOR = Blue Force, Bn = Battalion, TF = Task Force, OPORD = Operating Orders, FRAGO = Fragmentary Orders

Evidence on the workload required for more formal evaluations of command and control performance is available in almost every documented effort. For example, the relatively consummate evaluation of command and control information exchanges by Olmstead et al. (1973) required data reduction and analysis tasks that reportedly took months. Recall the database included all written and oral communications among the command and staff personnel. First, each analog communication had to be manually transcribed and sorted into predefined evaluation categories and events. Then evaluators collated all communications related to a particular "probe" event into a consolidated probe record. Subsequent analysis identified the organizational process (e.g., sense, communicate, decide) performed by each information exchange and assigned a performance competence score. Based on these component scores, composite process measures and scores, as well as overall competence scores, were developed. While analog command and control media account for only part of this evaluation's workload, it is a part clearly evidenced in Crumley's dire conclusion: "The data collection process, transcription of recorded oral communications, and extended analysis process do not lend themselves to routine application beyond the laboratory world" (Crumley, 1989, p. 20).

The impact of analog media is also evident in the ACCES work previously described (Halpin, 1992). As overview, recall that the overall volume of detailed information associated with command and control measurement resulted in the need for primary measures to avoid getting lost in the details. The burden associated with analog media is especially clear in the ACCES data collection procedures. These included the requirement to record all relevant information exchanges in journals, record key information from posted map overlays and status boards, and record all critical tactical and training events throughout the exercise to maintain the overall context of the mission.

The subsequent ACCES data reduction and collation processes required manual integration of the data from each collector to connect data points collected across all components of command and control as well as battlefield context into a meaningful pattern or finding. A notable example was the analytic process described to assess decisions on force withdrawal. This example underscored the breadth and depth work required for more thorough evaluation of command and control performance. The limitations of analog media contribute to this workload and require human evaluators to fully document and assemble data on command and control performance and the context in which it occurs.

Overall, the problem of manually burdened measures due to analog media limitations severely compromises most evaluations of command and control performance. This burdened measurement problem is not unique to the Army. Test and evaluation efforts in the Air Force, for example, incur more people, data, and processing than ever before (Gawron, 2000). And as the complexity of systems and missions increase, test and evaluation becomes more difficult and time consuming. To address this problem, the Air Force has initiated the development of a computer-based and highly integrated evaluation system called the Test Planning, Analysis and Evaluation System (Test PAES). A key design goal of this emerging system is to provide an integrated multimedia data analysis capability. The evaluators interviewed during the initial development stage of this system underscore the need for digital evaluation solutions to analog media limitations: " 'Do you know how many hours I waste matching strip charts to timelines to video to audio?' was the unanimous complaint of experienced test specialists" (Gawron, 2000, p. 24).

Summary of Background

This Background chapter began an examination of Army's evaluation requirement and evaluation methods as documented in foundation training documents. Contemporary factors undermining the Army's emphasis on evaluation were identified, including personnel turnover and warfighting complexity. In particular, shortcomings in the area of command and control training evaluation were highlighted. Proposed solutions to evaluation inadequacies addressed the need for more effective and efficient methods for training and evaluation. Proposed solutions also urged a renewed commitment to traditional Army evaluation methods, including reinstatement of traditional task, condition, and standard performance-based training.

A brief review of the behavioral science literature for evaluative methods applicable to the Army's emerging evaluation requirements was not very productive. Overall, members of this community acknowledge a major need for improvement in the means for performance assessment. More specific issues identified included the fact that training and performance evaluation methods were far too often piecemeal, irrelevant, and subjective. Moreover, the cognitive and collaborative nature of computer-mediated work complicates evaluation, especially in job domains where the boundaries and organizational structure are less concretely defined. Moreover, as the intellectual basis of work expands, more traditional task analytic methods to specify job requirements are proving inadequate. Proposed solutions included the need to apply digital technologies to help with the evaluation challenges they create. The ability of digital systems to represent information in graphical or pictorial format may provide visualization tools that help both users and evaluators. Similarities in cognitive and computer functions suggest the need for evaluation methods that identify and specify the behavioral and cognitive protocols used in performing job tasks and functions.

This Background also concluded that adequate evaluation of conventional command and control training and performance is challenging and rare. From a military perspective, a primary inadequacy in command and control performance evaluation is an inability to account for more of the variance associated with mission outcomes. The complexity and confusion inherent to warfighting, however, and situation dependent nature of mission outcomes complicate prediction. Moreover, the complexity and interrelation of command and control components and the limitations of analog media complicate measurement. Concerning command and control training evaluation, a primary inadequacy is the inability to provide more timely and meaningful training feedback.

From a research perspective, a primary inadequacy in command and control evaluation literature is a failure to routinely employ reliable and objective methods and measures. Although more objective methods and measures are available (e.g., Halpin, 1992; Olmstead et al., 1973), they are rarely applied. A primary reason identified for this failure was the heavy human workload associated with collecting, reducing, analyzing, and integrating conventional command and control data. A contributing factor to this workload is the passive and disjunctive nature of the media used in conventional command and control that require intensive human evaluation resources. As a result, conventional command and control evaluation is often inadequate and characterized by an over reliance on subjective measurement methods.

This Background chapter examined the Army requirement for more effective and efficient evaluation methods and measures, particularly in the area of conventional command and control. The Army's transition to digital technologies, such as instrumented command and control systems and military simulations, poses even greater challenges to training and performance evaluation requirements. This transition also affords opportunity. The Findings chapter examines and illustrates how digital evaluation opportunities can and should help solve the Army's training, and particularly, evaluation requirements.

FINDINGS

Challenges to training and evaluation intensify and expand as the Army moves toward an information age force empowered by digital technologies and, in particular, C^4I systems. This chapter begins by examining some ongoing and overarching efforts by the Army to identify and address the training and evaluation challenges that confront current and future forces. A central theme within this broad context of work is that training and evaluation researchers must help the Army determine how to apply digital technologies to meet these challenges. In support of this work, this chapter advocates the use of digital technologies, including digital instrumentation and the use of automated performance measures, for more effective and efficient methods for evaluating command and control training and performance.

This Findings chapter examines how the Army might apply digital technologies to evaluation with a focus on command and control performance. We begin with a brief overview of how digital systems, especially C^4I systems, are impacting training and evaluation in the Army. Attention then shifts to the focal issue of digital command and control evaluation. The heart of this chapter is titled Digital Command and Control Evaluation and it is divided into three topical categories: Challenges, Opportunities, and Examples.

Despite the purported potential of digital measurement methods, much of that potential is not yet realized, perhaps, not even envisioned. This chapter, therefore, includes a number of prototype concepts for command and control methods and measures that require further research and development. The examples are organized under three key evaluation issues that address the potential of digital technologies and the requirement for additional research and development in order to realize that potential: data integration, data mining, and data visualization.

Digital Training and Evaluation in the Army

The Army's research and training communities are working hard to design, develop, and implement more effective and efficient training and evaluation strategies and methods. A major impetus for this work is the challenge and opportunity posed by the Army's investment in digital technologies. By design, TRADOC leads and coordinates these extensive efforts. Moreover, TRADOC's guidance on training development stresses the need to leverage digital technologies to support training and evaluation efforts (U.S. Department of the Army, 1995).

Key elements of this guidance were provided in TRADOC's overarching Digital Learning Strategy (U.S. Department of the Army, 1998a). This evolving Digital Learning Strategy is expected to encompass the Army's digital training requirements at individual, small group, and collective levels. Notable aspects of this strategy, for present purpose, are its emphasis on evaluation to ensure proficiency and its explicit expectation that digital technology will support evaluation requirements. As this strategy develops, digital methods for embedding performance assessment must be designed and developed.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) in concert with TRADOC is developing a comprehensive research plan to address digital skill training and retention issues (Moses, in preparation). This plan will outline ARI's program of research designed to address the Army's most pressing needs in the area of digital skills and learning, and its consequent benefits to the Army. Concerning evaluation, a major need identified in this plan is the requirement for evaluative methods that certify digital skills and provide training assessment support for trainers. Evaluative needs identified in this research plan include: measurement and feedback for progressive skill levels; recording and transmitting data; competency tests, skill certification, and licenses; and obtaining external, objective observers.

As part of ARI's research on digital training and evaluation, the Future Battlefield Conditions team at Fort Knox designs and develops prototype training and evaluation methods. This team's research objectives address future staff training and evaluation requirements. More recent team efforts have focused on the design of team training and assessment strategies, in a digital environment, using current research in the cognitive and behavioral sciences. The training setting for the team's work is the future digital tactical operations center at brigade and below, and the research environment leverages advanced digital technologies, such as instrumented C^4I systems and virtual simulation. The team's ongoing work, including this report, focuses on the requirement for automated performance assessment methods for command and control on the digitized battlefield.

In addition, an informative analysis on the overall scope of the training challenge for a digitized force is provided by Dierksmeier et al. (1999). This analysis describes how the Army is addressing and might further address those challenges, including the use of structured simulation-based training in the Army's Close Combat Tactical Trainer-Digital (CCTT-D). The lessons learned from this analysis are summarized in two major themes: the need to integrate training requirements and digital technologies; and the need for comprehensive Training Support Packages (TSPs) for digitally equipped units. Consistent with the focus of this report on the need to apply digital technologies to evaluation and the lessons noted above Dierksmeier et al. concluded: "...a key recommendation offered in this report is the development of a multi-functional automated performance measurement database that complies with the Army's latest simulation technologies..., supporting both training and research objectives for Force XXI" (Dierksmeier et al., 1999, p. 3).

Digital Command and Control Evaluation

With the advent of digital command and control systems, many conventional evaluation challenges remain and a host of new evaluation challenges emerge. This section briefly reviews how some of the previously described challenges complicating conventional command and control evaluation also confront digital command and control evaluation. In addition, this section reviews some new evaluation challenges introduced by digital technologies and by contemporary requirements for more diverse and complex missions.

Next, the central theme of this section examines the opportunities afforded by digital technologies for command and control evaluation. From the outset, these digital opportunities are only partial solutions to only some of the myriad challenges facing command and control evaluation. Perhaps, the most difficult challenges are inherently subjective, such as determining the value of a command decision or mission outcome that entails human lives. More objective and robust data on command and control performance from digital technologies, however, should provide an improved basis for making decisions, for obtaining desired end states, and for evaluation.

Challenges

Perhaps, the oldest and most basic challenge facing any new system is proving its value. Unfortunately, performance evaluations of the Army's digital command and control systems have failed to do so. Across a series of evaluations called Army Warfighting Experiments (AWEs), including force-on-force exercises at the NTC, Army evaluators "...concluded they could not identify any significant increase in force effectiveness over baseline units" (General Accounting Office, 1999, p. 14). The force effectiveness measures used for most AWEs included lethality, survivability, and operational tempo.

While explanations of such failures exceed the scope of this report, such conclusions reinforce how evaluations of digital command and control systems face the same mission outcome challenge as analog systems. Another recurrent challenge noted by the GAO is the system-of-systems complexity inherent to command and control evaluation. Their report raises concerns about the performance uncertainties associated with the 16 high-priority subsystems or components that currently comprise the overall command and control system for the first digitized division. It also stresses that incompatibilities between these subsystems must be resolved before an integrated, digital command and control system can be successfully fielded.

Moreover, the challenge of burdened methods and measures for command and control evaluation remains and, in fact, digital C^4I systems may actually increase that burden. A recent analysis of live force training requirements on the digital battlefield noted that C^4I systems increase the training analyst's load, and force a faster feedback process to provide meaningful training experiences (U.S. Army Training Support Center, 1996). In particular, this analysis stressed that training *feedback overload* must be avoided for C^4I-based training. The analysis of OC and analyst workload by Brown et al. (1998) also proposed that digital command and control will almost inevitably increase evaluation workload. Predictions of greater workload are based on their observations that digital systems increase the volume of information exchanged, decrease eavesdropping capabilities, and require over-the-shoulder observation of human-computer interaction. However, the prediction that digital systems invariably increase evaluator workload is *countered* in a subsequent section of this report titled Less Burdened Measurement Methods.

In addition, new evaluation challenges are introduced by digital technologies and by contemporary requirements for more diverse and complex missions. Alberts et al. (2000) provides an overview of some of major new command and control challenges including:

- Multitude of New Mission Changes
- New Role of Command and Control
- Recognition of the Importance of Coevolution
- New Architectural Forms
- From Uncertainty Reduction to Better Battlespace Awareness

Interested readers are referred to Alberts' extended discussion of these challenges and how they might impact both the conduct and evaluation of command and control. The following paragraphs concentrate on a few of the major challenges identified by Alberts et al. to clarify their potential impact on evaluation.

Alberts et al. suggests that the most profound challenge may be that digital technologies will result in a new role for command and control by transforming it from a force multiplier to an identifiable force. The integration of C^4I information systems and weapon systems will result in attacks by and on information systems that may equate to decisive engagement. On the future battlefield, information systems will directly enable commanders to defeat adversaries and reduce friendly force casualties. To the extent information becomes a force multiplier, the indirect relationship between command and control and mission outcomes (e.g., lethality and survivability) will become much more direct. The notion of digital information as a force reflects the conjunctive nature of digital media that directly couples command and control systems to weapon systems. Such coupling can transform a digital message on target location to a weapon system's automatic slew and fire to that target.

From an evaluation perspective, an even greater challenge posed by digital technologies is their rapid and pervasive impact across doctrine, training, leadership, organization, materiel, and soldier (DTLOMS) domains. These changes affect performance and training at individual and collective levels. They also affect relationships among organizations and workers, such as commanders and their subordinates, and even the nature of military operations (Alberts et al., 2000). The potential impact of digitization is often referred to as a revolution in military affairs. By design, the conjunctive nature of digital technologies is extending the connectivity and interdependence of digital command and control systems to weapon and support systems. The iterative and interactive process of change, referred to as spiral development, induced by digital technologies creates co-evolutionary requirements across interdependent systems and domains.

Failures in meeting this expanding spiral of requirements may partially account for the evaluation failures reported for the AWEs. Evaluators have often found that the introduction of technology results in little or no improvement in an organization's performance (Nickerson, 1995). Efforts to explain this effect, often characterized as the "productivity paradox," underscore the point that investments to insert technology require even greater investments in new organizational structures and processes, and particularly training (Alberts et al., 2000). This is especially true when evaluating gains from digital information technologies that markedly impact the entire organization and how it conducts business. An analysis by Brynjolfsson and

Hitt (1998) suggests that costs for these "other" investments may exceed hardware costs by tenfold, and that the time lag before productivity gains are realized from the technology is typically 4-5 years. Their analysis discloses, moreover, that unless an organization makes these other investments, it may be *worse off* after a technology insertion.

Even after adequate investments in technology and organizational change are initiated and maintained, evaluation strategies and methods must repeatedly adapt to the ongoing process of change induced by digital technologies. Strategies for adapting evaluation methods to changing environments, however, are emerging. For example, a comprehensive strategy for evaluating the impact of complex information technologies is provided by Pejtersen and Rasmussen (1997). This strategy reflects the iterative effect of technology advances on established practices and traditions within an organization. Their characterization of this situation includes the following (Pejtersen & Rasmussen, 1997).

- The rapid pace of technological development makes a smooth empirical and incremental development of systems difficult.
- New technology upsets the traditions and practices that encourage mutual understanding between designers and users or consumers during periods of technological stability.
- The new *multimedia technology gives new means for recording and analyzing data*, a development that has resulted in a great number of usability laboratories in industry.

The strategy proposed by Pejtersen and Rasmussen (1997) encompasses a broad range of evaluation objectives and provides a framework for partitioning this range based on six different work domain boundaries. The inner boundary, for example, focuses on user characteristics aligned with laboratory experiments directed at comparing whether the system's information presentations match users' sensory and motor characteristics. The outer boundary focuses on actual operational characteristics of the system aligned with field studies in actual work environments directed at comparing whether the system characteristics match users' operational requirements.

Overall, the evaluation challenges associated with digital command and control are formidable. Organizational changes especially in command and control doctrine and training require commensurate changes in evaluation methods and measures developed for conventional organizations. At a minimum, evaluation methods and measures must be converted or modified to reflect changes in tasks and standards. Similar method and measure modifications are required to address changes in warfighting conditions, including more diverse and complex missions. To address justification challenges about productivity improvements, for example, evaluators might account for "other" investments and time, and employ broader and longitudinal measurement methods. At best, digital technologies may afford unique opportunities for meeting these challenges and developing more effective and efficient evaluation methods and measures.

Opportunities

For electricity not only gives primacy to process, whether in making or learning, but it makes independent the source...from the location of the process. (McLuhan, 1964, p. 346)

The Army's investment in digital technologies assumes that the opportunities afforded by advanced information systems will maintain a dominant force. At a visionary level, this is a force of cyber warriors with humans and computers allied as a joint cognitive system (Cook & Woods, 1996). Digitizing the battlefield is defined as: the application of technology to acquire, exchange, and employ timely information horizontally and vertically integrated to create a common picture of the battlefield from soldier to commander (U.S. Department of the Army, 1998b, p. 5). The primary technology in this modernization effort is the C^4I system that serves as the "tool" required to perform the command and control functions directed at mission accomplishment. One way of epitomizing the operational opportunity provided by C^4I systems is the expectation they will provide all combatants and supporters a common picture of the battlefield situation.

The purpose of this section is to examine the opportunities afforded by digital technologies for evaluating training and performance, particularly in the area of command and control. This examination is organized under five topical categories or sections. First, Digital Media considers how the very nature of digital technologies supports interactive and collaborative information exchange and processing. Next, Digital Instrumentation considers the types of command and control data that might be automatically collected and processed by digital technologies, particularly instrumented C^4I systems in concert with military simulations.

Then, the section titled Increased Scope and Precision examines how digital technologies allow evaluators to collect more exact measures of performance and adjust the range and selection of data to include any or all C^4I users at any or all times during a mission. The section titled Meaningful Evaluation then considers how digital technologies might enable users and evaluators to visualize command and control process and product data in its performance context. Finally, the section titled Less Burdened Methods and Measures examines how digital technologies might reduce the high workload associated with command and control evaluation.

Digital Media

The term digital media describes a broad range of electronic information systems that process and exchange data discretely represented by numbers in binary format. The uniform binary format shared by digital technologies has resulted in the convergence of computers and media into multimedia (Chignell & Waterworth, 1997). In this report, the term digital media is used to reflect this convergence and provide a more direct contrast with the analog media used to conduct and evaluate conventional command and control.

Analog media were previously described as passive and disjunctive "carriers" of information. Limitations imposed by analog media on the conduct and on the evaluation of command and control performance were described. In particular, these descriptions focused on the increased human workload required by performers and observers in order to process and communicate command and control information.

In contrast, digital media are characterized here as interactive and conjunctive. Interactive reflects the ability of digital technology to accept user input and respond with information conditionally and non-linearly (Locatis, 2000). Conjunctive reflects the ability of digital technology to join information within and across media. It is a union made possible by the general standard imposed by digital media, a uniform binary format of 0's and 1's.

This report proposes that the defining and most salient aspects of media are their ability to act on and access information. Figure 1 provides a notional comparison of analog and digital media based on their information processing capabilities. This figure depicts a continuum of capability for acting on information (passive-interactive) and accessing information (disjunctive-conjunctive).

This report's approach to distinguishing media capabilities relates to what Kozma (1994) and others refer to as the underlying structure and causal mechanisms by which media enable information processing. Media are often classified at the surface level of symbol systems and stimulus categories including text, numbers, voice, and still versus moving pictures. At a more fundamental level, however, media can be distinguished by their ability to *process* available and diverse symbol sets into information and meaningful representation. Kozma asserts that digital media create a unique job or training environment that supports information exchanges and interactions between coprocessors, the user and the computer. At the surface level media transmit data to users; at a more basic level they construct information and knowledge *with* and for users.

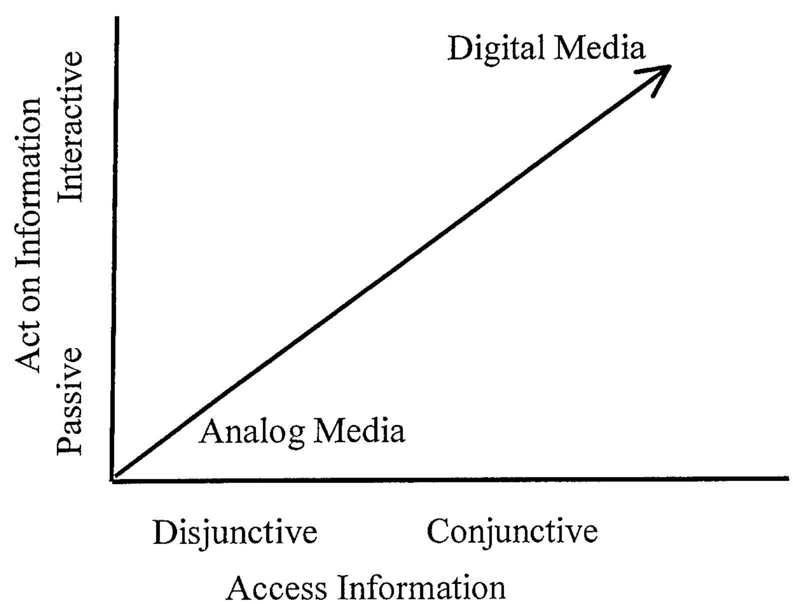


Figure 1. Comparison of analog and digital media capabilities for acting on and accessing information.

As information processors, digital media interact with the information they carry. Digital media are also designed to interact with and support human users, including senders and receivers. To facilitate interaction, digital messages usually include not only message data, but also metadata on how message data should or might be processed. This metadata can provide contextual information about how the message data relates to other data, or how it relates to a broader domain of information. For example, digital image files often include ancillary data on image size, exposure date and time, flash used, exposure settings, compression schemes and data formats. Digital media not only transmit data but also embed it in form of stored instructions or programs for interacting with the message data. Based on experience with message data and embedded data, a digital processor can even alter its instructional set or program to improve interaction.

The digital processors in computers and related digital media emulate and automate many of the logical computations or functions normally performed by humans. These functions filter, sort, collate, organize, compare, and transform data. In essence these functions convert data into information, such as spatial patterns and time sequences. For example, C^4I processors can transform stored data, such as terrain, and command and control data, such as reports and overlays, into a dynamic picture of the battlefield situation.

As exchangers of information, digital media are conjunctive. They readily facilitate the collaborative exchange of information within a medium and across media, as well as with other technologies, such as digital simulations and weapon systems. Moreover, they connect individuals into groups, societies, nations and global networks. Protocol data, for example, is explicitly designed to maximize information exchange with other media and systems. Data embedded in hypermedia provides links to related data across diverse media formats. In fact, the term "hyper" media underscores a capability that exceeds the limitations inherent to most media, particularly analog media. A defining aspect of hypermedia, and even hypertext, is the ability to access information beyond the immediate informational resource or domain. It is the conjunctive nature of digital media that provides ready links or connections to additional data and information within a medium (e.g., hypertext) and across media (hypermedia).

A common and compelling example of how digital media link data and information are web-based search engines. The ability of such search engines to access requested information is unprecedented, even with current technologies and databases. During the drafting of this report, for example, a search on the term "data logging" by the search engine resulted in the following: "Google results **1-10** of about **549,000** for **data logging**. Search took **0.12** seconds." (Google, 2000).

Over a half million "hits" on a topic like data logging is not only a lot of access to information requested, it is fast and direct access. In the reported interval of .12 seconds, literally faster than an eye blink, a requester is provided not only a stand-alone list of references, but also a direct and immediate connection or hyperlink to each source found in the search. Many powerful search engines or tools for accessing information are freely provided to any web-based user. Moreover, the ability of digital media to access information will only improve. Improvements will include technical innovations such as improved software routines and algorithms, as well as faster and more capable hardware for searching and channeling information to users. Improvements will also include the ongoing extension of digital media into public and private domains, the evolution of literally global and beyond information networks and their merger into one network.

Each segment of a digital message, often called a packet, contains information on where it came from, where it is going, how many steps or network nodes it has passed on the way, and how to reassemble itself and related segments into the full original message. In collaboration with other digital technologies, digital media can also reassemble the message context. An example might be the ability to play back a recorded military simulation exercise that recreates battlefield conditions at the time the message was sent. Moreover, collaboration across a unit's C^4I systems, can convert stored and exchanged data to portray a common picture of the battlefield situation.

The uniform binary format of digital information not only links different types of digital media it also supports data correlation and synchronization enabling multimedia, and data transformation enabling mixed media. In contrast, analog media are generally fixed-function providers dedicated to a single form of data—text, sound, graphic, or image. Printed books have evolved to include graphs and picture images, and motion pictures merge aural and visual data forms, but they do not "talk" to each other.

In contrast, digital media integrate multiple data formats into a conjunctive multimedia form. Digital media not only convey multiple forms of data, they allow users to access the data in nonlinear and random associations, and to manipulate and rearrange the data and data structures into new forms and patterns. While multimedia capabilities are increasingly common, the potential of mixed media capabilities is only emerging. With mixed media, sometimes called synaesthetic media, information normally unique to one media and sensory mode, such as a piano performance recording, can be transformed into a different media and a different sensory mode, such as a printed score, or vice versa (Chignell & Waterworth, 1997).

The information processing and exchange capabilities of digital media have transformed how workers and organizations work, including the area of command and control (Woods, 1996). For example, the challenges cited by Alberts et al. (2000) included transformation of the work and function of command and control from a force multiplier to an identifiable force. More specifically digital media have automated or partially automated many of the human information processing and exchange functions essential to command and control. In doing so, digital media have increased the interdependence between humans and machine. Digital media are not stand-alone devices. Unlike analog media, digital media are becoming integral to command and control performance. The following assessment makes this clear: "We need to think of new automation as part of this control and management system rather than simply divide the world into machine and human parts" (Sarter, Woods & Billings, 1997, p. 1938).

From an evaluation perspective, the characteristics of digital media and particularly their integral role in work performance provide unique and powerful opportunities for collecting and analyzing command and control data. Prior discussion of analog media focused on how media limitations increased the workload associated with nearly every phase of evaluating command and control training and performance. For example, radio voice data require that OCs and analysts transcribe voice messages on note pads, observation sheets or maps during the exercise. Moreover, tracking and integrating a series of related voice communications, even relays of the same message, is a tedious and complicated task. In contrast, content data from digital messages and reports can be automatically stored and retrieved. Metadata such as sender and receiver identities and tracking information on how many steps or network nodes a message has passed are also automatically stored and retrievable.

Digital versions of maps, orders, and overlays can also reduce evaluation workload. These digital formats should greatly reduce the need for evaluators to manually collect and transmit these products across numerous physical locations. Digital processors can automatically check on the internal consistency of similar products. In particular, the conjunctive nature of digital media can automatically integrate information, such as overlays and battlefield reports, which had to be manually integrated with analog media. For example, enemy information in the form of digital Contact and Spot reports along with operational overlays can be automatically transformed into graphical formats and posted on an evaluator's display of the battlefield situation.

Most importantly, digital media can largely automate the tedious and error-prone work of relating command and control performance to the tactical context in which it occurs. This is particularly true when the exercise is conducted in a digital warfighting simulation that is fully recorded. However, even the partially instrumented ranges used at the CTCs for live training record a substantial portion of the tactical context. The conjunctive nature of digital media readily supports the evaluation requirement to correlate task performance to task conditions and provide an integrated representation of both on a digital display.

Overall, digital media are expected to assist evaluation efforts by automating much of the workload required to perform nearly every step in the evaluation process. This includes key aspects of analysis, such as pattern and trend recognition. Evidence on the potential reduction in evaluation workload due to digital media will be considered under the section titled Less Burdened Methods and Measures. Subsequent sections will also consider how digital media might increase the quality and quantity of data available to evaluators and provide more meaningful information. The more immediate concern addressed in the following section, however, is to more specifically identify the types of performance data available from digital media that might support evaluations of command and control performance and training.

Digital Instrumentation

Digital technologies are uniquely suited to automatically collect user performance data. In fact, most computers can and do routinely log or track all user inputs and system responses. When personal computers were actually "personal" and more disjunctive, not networked to other computers, digital logs primarily served the user. Common examples of how such computer logs aid user access to information, sometimes called navigation, include a "Back" key and a list of most recently opened files.

An example of how computer logs aid user interaction with information is the "Undo" command or key in Microsoft Word™[1] applications. In fact, a ready example of a digital tracking log available to most computer users is the list of undo actions accessed by this command. This list identifies all basic user interface actions, such as mouse selections or keyboard inputs, during the current session. This list or log may entail hundreds of user actions or interactions, depending upon the frequency of actions and duration of the session. A related example is the "Track Changes" function that allows a user to compare changes made to a current document version against an earlier version. Illustrative examples of Undo Action and Track Changes functionality are provided in Figure 2.

As digital media become more networked, the information they process including digital logs becomes less private but more useful to others, including evaluators. More basic techniques for disseminating log data simply capitalize on the internal tracking routinely performed by computers. Employee monitoring software, for example, typically reports a system's internal log processing data by routinely sending discreet emails to another computer or server. The content of these email messages identifies page content and employee interactions and allows employers to "monitor" employee monitors in real-time or later. The relatively minor cost for such applications, frequently less than $100 currently, underscores how readily employers, trainers, and evaluators might obtain digital tracking data on performance.

More sophisticated techniques for collecting log data are continuously being developed and refined. For example, one of the more infamous instances of digital tracking is a cyber cookie. A cookie is a "crumb" of software stored in the computer of anyone who visits an Internet website. Websites installs these cookies on the visitor's hard drive to track the visitor's behavior at the site to obtain marketing information, for example. Such cookies often store the last date and time a user visited a site. This information is used for multiple purposes, such as, customizing the information available, ensuring "new" information is presented, and collecting demographic information provided by the user.

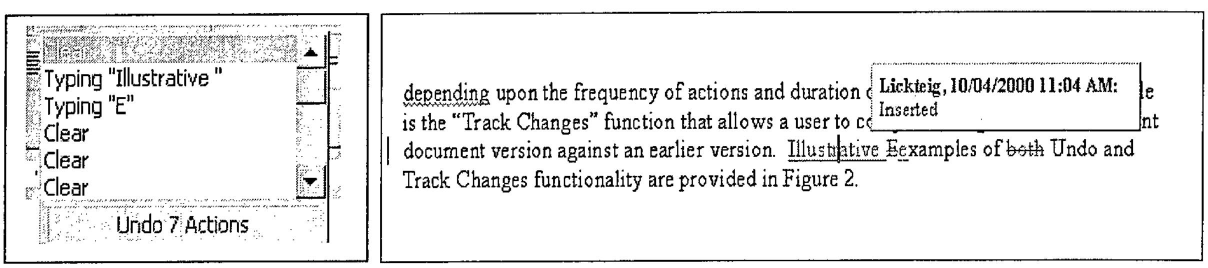


Figure 2. Digital tracking examples of Undo Actions and Track Changes provided by Microsoft Word.™

Another popular instrumentation technique is a user identifier, a signature-type serial number unique to each computer or media component that is automatically inserted into outgoing communications, such as electronic mail. On one hand, digital system providers have reportedly installed this capability on mass-marketed software products without informing

[1] Microsoft Word is a registered trademark of Microsoft.

buyers or users. On the other, Internet standards organizations are contemplating making such instrumentation a protocol requirement to guide information exchange.

The evaluative potential of digital systems may be most commonly associated with computer-based training and testing. Over a half million individuals completed licensing and admission exams on computers in the 1997-1998 academic year and many of these tests included open-ended items such as essays (Bennet & Bejar, 1998). While earlier versions of computer-based tests relied primarily on traditional multiple-choice items, subsequent testing methods have capitalized on the ability of computers to track problem solving interactions to expand the range of test items. For example, when the evaluative goal is to assess process-oriented skills, such as math reasoning, automated scoring routines capture interactive evidence of solution processes and not just solutions.

The article by Bennet and Bejar (1998) examines problem construction, automated scoring algorithms and validity issues across a range of test performance conditions that include both constrained and naturalistic problem solving. The authors' extensive experience with digital tracking techniques at the Educational Testing Service results in a balanced and broad assessment of the pitfalls and potential of computer-based evaluation. Pitfalls, for example, include their cautionary reminder that evaluators avoid thinking that just because some surface aspect of performance can be automatically measured it should be; it may not reflect the target construct. Overall, however, they conclude that computer-based testing provides unique opportunities to improve performance assessment, even for complex and unconstrained tasks, and to reduce the bias and burden associated with human evaluators.

Despite the alleged potential of digital technologies for performance measurement and evaluation just reviewed, that potential is largely unrealized. Although computer-based work and training mushrooms, the number of researchers who have attempted to explore this technology is surprisingly small (Burroughs et al., 1999). Why? The report by Burroughs et al. cites several factors including: although computers are now ubiquitous, they are only recently affordable; development of the measurement methods is notoriously long; and, the lack of methodological guides has greatly increased the difficulty of exploring new and innovative assessments. In particular, that report stresses an approach to evaluation that judiciously allocates measures to humans and machines based on the things they do best.

After a decade of research at the Wilson Learning Center, a telling conclusion concerning computer-based performance measurement probably still applies: "At the time, computerization seemed liked the ultimate panacea. It was, and was not" (Burroughs et al., 1999, p. 223). This balanced perspective is based on an extensive body of work at the Wilson Learning Center that successfully applied digital technologies to assess organizational behavioral skills. For example, Burroughs et al. reports that digital methods and measures of employee skills reduced the time required by human assessors by almost 50 percent. For the Department of Defense efforts in selection and placement, Burroughs et al. states that computerized enlistment testing saves approximately five million dollars a year. In addition, computer-based enlistment testing allows for portable and/or walk-in administration and saves many person hours by eliminating the need for recruiters to accompany test applicants to conventional test sites.

Our growing reliance on computer-mediated work underscores the potential of digital technologies to automatically collect and analyze real world performance data. This is particularly true of work conducted on microworld systems in which inputs and outputs correspond to movements in the physical world. The Army's C^4I systems are real world systems intended to depict a microworld representation of the battlefield situation and surrounding environment in a manner that enables actions and thought on that environment, at the level the user chooses (Rasmussen & Pejtersen, 1995).

As C^4I systems become more integral to the performance of individual and collective tasks, the tractable human-computer interactions associated with these systems become more critical to evaluation. The more a C^4I interface becomes the primary point, or means, of interaction between soldiers, the more instrumented data on human-computer interactions represent essential process and product metrics for command and control evaluation. The more command and control evolves from a force multiplier to a force, the more instrumented systems can provide automated and objective measures of performance, effectiveness, and their inter-relationship.

However, an essential requirement for achieving the evaluative potential of digital technologies is the instrumentation of these technologies, particularly C^4I systems. The importance of instrumentation is strongly underscored in the lessons learned by the Information Systems Engineering Command and its lead laboratory for systems testing, the Technology Integration Center (Ward, 2000). As this Center moved from what they called "single-box" evaluations to ensuring systems integration, its evaluators wrestled with costly and disruptive methods such as installing parallel systems that emulated the actual systems being tested in the field.

Instrumentation proved a more efficient and effective method for evaluating system integration, as stated by the Information Systems Engineering Command's Senior Technical Director, Dr. Michael Gentry: "The real breakthrough came when we were able to capture all the keystrokes being used in the field....For the first time we could see where the bottlenecks were and recommend hardware and software fixes" (Ward, 2000, p. 45). Similarly, this report promotes a C^4I instrumentation requirement for more efficient and effective methods of evaluating *soldier*-system integration, especially in the area of command and control performance.

Instrumentation of a digital information system should result in a log of all soldier-computer interactions performed on that system. Each instance of a system, namely each operator's C^4I device, should maintain a unique log. As individual C^4I devices are networked together, the log would include information exchanges with other C^4I systems, such as messages received and transmitted. The conjunctive nature of digital media, moreover, readily expands this network to include information exchanges with other types of digital systems including sensor systems and military simulations, as previously discussed.

The definition of C^4I system instrumentation proposed in this report, reflects the potential coupling of digital media: *instrumentation equates to a log of all soldier-computer interactions correlated with the battlefield situation in which they occur.* Unfortunately, no Army C^4I

systems are fully instrumented, and most are not instrumented at all (Brown et al., 1998). The following sections examine how digital technologies, particularly C^4I instrumentation, might provide: more balanced measurement methods, increased measurement scope and precision, more meaningful evaluation, and less burdened measurement methods.

More Balanced Measurement Methods

Digital evaluation technologies provide direct measures *of* performance. In contrast, evaluations of training and performance, as reviewed, disclose a biased reliance on measures *about* performance, particularly subjective and imprecise measures. The value of direct performance measures, and a distinction between measures *of* versus *about* performance was aptly summarized by Prosser: "...but there is no [wo]man who would not accept dog tracks in the mud against the sworn testimony of a hundred eye-witnesses that no dog had passed by" (Prosser, 1964, p. 216). This section proposes that the ability of digital technologies to automate performance measurement can and should help balance measurement methods for evaluations of command and control performance.

Clearly, automated measures from instrumented C^4I systems are not a cure-all. As stressed, more traditional measurement methods are still needed, such as observation, interviews, and questionnaires. However, many of the perceived shortcomings in the use of more traditional measurement methods can be offset by the application of digital measurement methods. Overall, balanced measurement methods are needed that include both traditional and digital measurement approaches. To correct a perceived imbalance, however, this report focuses on the application of digital measurement methods to improve command and control performance and evaluation.

Evaluators from behavioral science and the military extol a multimethod measurement approach (e.g., Cannon-Bowers & Salas, 1997). For example, the seminal work by Webb, Campbell, Schwartz and Sechrest (1973) stresses: "No research method is without bias or weakness" (p. 1). To overcome the weakness associated with any measurement method they stress that no method or measure should be used alone. Ideally, evaluations should include multiple methods of measurement such as observation, interviews, questionnaires, and direct measures of performance. Webb et al. note that once a proposition or hypothesis is confirmed by two or more independent measures, the uncertainty of interpretation is greatly reduced. The most persuasive evidence comes through a triangulation of measurement processes, in which three or more methods are used to measure the same thing. Regrettably, evaluators in the areas of training and social science tend to rely on a singular and subjective measurement method, namely the interview or questionnaire (Nickerson, 1995; Webb et al., 1973).

Similarly, Bornman's (1993) handbook on command and control measurement, for example, provides a lucid explanation and examples of why and how both subjective and objective measures are essential to military evaluation. Descriptions and examples of traditional types of subjective and objective measures commonly used to evaluate military operations, including command and control performance, are provided in Table 2. As Bornman explains, military evaluators often deal with conceptual and perceptual concepts, such as desired end states and mental pictures rather than physical entities/realities. Because such concepts are judgmental and based on multivariate inputs, it is much easier and in some cases essential to evaluate them in a subjective manner. Mission objectives and military decisions, for example, often entail the value of lives relative to the social/political values for which those lives are risked.

Measures of Effectiveness (MOEs) address the operational capabilities of a force, including the aforementioned battle outcomes, and are the "bottom line" for military evaluation. However, MOEs are inherently subjective. Primarily, because they are dependent upon the scenario or mission in which they are collected. Bornman notes that the unique and multivariate scenario/mission factors that determine outcome are not replicable and, therefore, MOEs are not reliable. The subjective nature of MOEs also applies to measures of system effectiveness, such as the command and control MOEs listed in Table 2.

Table 2

Measure Types and Categories Used in Military Evaluations*

Measure	Measure Description
Objective	Objective measures are based on facts and tend to be unbiased. They are largely captured through physical observation or recording of some physical change observed about physical entities (e.g., time, shape, physical condition, quantity).[1]
Subjective	Subjective measures place emphasis or reliance on one's own moods, attitudes, opinions, experiences or values. They may be biased in many ways and forms. They reflect the perceptions of the observer or multivariate inputs. As both command and control measures of effectiveness and measures of force effectiveness are scenario dependent, they are by their very nature more in the subjective realm.[1]
Effectiveness	Measures of Effectiveness (MOEs) should be defined to measure operational capabilities in terms of engagement or battle outcome. Examples include: percent of threat tanks defeated, percent of threat aircraft defeated, percent of friendly weapon systems surviving an attack.[2]
Performance	Measures of Performance (MOPs) should relate to the measure of effectiveness such that the effect of change in the measure of performance can be related to a change in the measure of effectiveness. Examples include: probability of target detection, call completion rate, and time to process information.[2]
Command and Control Effectiveness	Command and control MOEs assess how a command and control system affects the other entities in an operational environment (e.g., reaction time, susceptibility to deception); measured relative to some perceived standard, often implicit (e.g., how a perfect command and control system would perform); are scenario dependent.[1]
Command and Control Performance	Command and control MOPs represent a measurement of an attribute of system behavior (e.g., throughput, error rate, process resource requirements such as time, space, and quantities of physical entities); relate to the inherent properties of the system and/or internal to the system being analyzed; are scenario independent.[1]

*Note. Based on edited excerpts from Bornman[1] (1993, pp. 2-3) and Kass[2] (1996, pp. III-4).

More objective Measures of Performance (MOPs) are needed, therefore, to establish a solid foundation for MOEs. Usually, MOPs are based on more physical entities that are directly measurable, observable, and independent of the scenario or mission in which they occur, as indicated in Table 2. Command and control MOPs, for example, deal with how system entities—soldiers and machines—are arranged and organized, and what they are doing or being used for.

We focus this report, therefore, on the need for more balanced and objective MOPs for command and control, on the need for direct measures of performance. Notably, observation-based measures are generally regarded as objective. However, the objectivity of observational data, as well as its reliability and validity, is highly dependent upon observer skills, workload, and data collection instruments.

Unfortunately, as previously considered, OCs responsible for training and evaluating command and control performance confront high workloads that "divert" the attention they can provide to careful observation and detailed data collection. In addition, the Training and Evaluation Outlines (T&EO) used by OCs frequently fail to specify many relevant dimensions of task performance, provide only cues to standard specification, and list tasks in strict chronological sequences that are unrealistic (Ensing & Knapp, 1995). Such limited data collection formats prevent the collection or capture of many meaningful performance parameters. However, recall that even more streamlined observer formats were recently developed to further minimize observer burdens (Hallmark et al., 1998). Observational data on command and control performance may be highly subjective with limited reliability and validity. And as more streamlined and global behavioral categories are used to encode observational data, the more observational data becomes an indirect and inadequate measure about performance.

In contrast, automated measures from instrumented C^4I systems provide direct measures of performance, particularly command and control performance. The more important C^4I systems become, the more soldier-computer interactions associated with these systems become critical and collectible. This report proposes that digital technologies, and automated measures in particular, are needed to help balance the overall set of measurement methods used to train and evaluate command and control performance.

Increased Scope and Precision

> *By instrumenting the electronic battlefield and issuing in an accessible, digital format, comprehensive and absolutely accurate data... networked simulation both enhances the most promising of our measures of collective behavior and makes them practicable.* (Fletcher, 1994, p. 268)

The ability of digital systems to track or log human-computer interactions provides an unprecedented opportunity for collecting and assessing more precise, or as characterized by Fletcher "absolutely accurate" measures of performance. This ability applies across a broad scope of human behavior, including psychomotor, procedural, cognitive, and collaborative performance.

Increased precision is not without cost or risk, such as missing the proverbial forest for the trees amid an overwhelming wealth of performance data. Increased scope compounds costs and risks, particularly in an age where masses of digital data are continually collected on global-size groups of individuals. However, digital measurement methods are rapidly improving our ability to use vast and detailed human performance databases. Digital measurement methods should greatly reduce cost, perhaps even risk, and establish bedrock performance data that greatly expands our understanding of both "the trees and the forest."

Psychomotor Skills. Psychomotor skills exemplify the need for and potential of digital technologies to provide more precise human performance measures. Psychomotor tests are relatively precise measures used to assess the psychophysical limitations of humans, and include simple reaction time, rotary pursuit, finger dexterity, and tapping and sorting speed. In general, they provide a "lean" measure of human processing speed and accuracy with the trappings of much cognitive processing removed. For example, correct responses are often known in advance. Formerly, psychomotor tests were widely used to predict individual differences in skilled performance for selection such as assembly and clerical workers, including over 600,000 individuals by the U.S. Army Air Force during World War II, as reviewed by Ackerman and Cianciolo (1999). However, their review stresses that psychomotor tests are being abandoned as impractical, despite their proven ability to predict individual differences in highly skilled performance, including the final phase of skill acquisition. Why the abandonment of psychomotor tests, why are they impractical?

Analog media, particularly the barriers associated with analog measurement methods, are the primary reason psychomotor testing is regarded as impractical. A summary of that reasoning and the analog measurement barriers identified and examined by Ackerman and Cianciolo (1999) is provided in Table 3. This table's descriptions of the barriers with analog device fabrication and calibration reflect the limitations of analog media, as previously discussed. The stand-alone or disjunctive nature of analog media requires individual design and fabrication, rather than a general purpose, multi-test, device. It also prevents media "crosswalk" for synchronization, including self-calibration and calibration across devices. Similarly, the barriers associated with psychomotor examiners, including training, monitoring and data collection, underscore the manual burden imposed by analog methods.

Digital measurement methods eliminate all of the analog barriers identified in Table 3, conclude Ackerman and Cianciolo (1999). That conclusion was based on two empirical evaluations of psychomotor abilities that entailed a detailed examination of construct and criterion validity over an array of psychomotor tests. Further, they state that digital technologies reduce additional barriers associated with analog testing, particularly the vexing problem of test or measurement revision. With analog devices, revisions to a test or measure often entail extensive and expensive hardware re-fabrication. With digital devices, revisions were easily and quickly accomplished with software updates via scripting programs, they report. In summary, they stress that digital testing devices readily render precise measures of psychomotor ability, and may enable a more complete assessment of fine psychomotor skills than ever before possible. Digital measures of psychomotor ability should improve performance prediction in dynamic jobs and "...make it feasible to fill in these important gaps in knowledge about the structure of human abilities" (Ackerman & Cianciolo, 1999, p. 270).

Table 3

Barriers Associated with Analog Psychomotor Testing Methods*

Barrier	Barrier Description
Fabrication	Each of the apparatus-based psychomotor tests had to be individually designed and fabricated, often to a high degree of precision, and at a high cost. There was no technological provision for a general purpose psychomotor testing platform.
Calibration	Psychomotor tests using specialized apparatuses require constant adjustment and calibration, to ensure that the variance in examinee responses is attributable to individual differences in ability, and not to differences across apparatuses.
Examiner Training	Use of specialized equipment requires a substantial amount of examiner training.
Examiner/ Examinee Ratio	…it was necessary to have one examiner for every four examinees, during psychomotor testing, to maintain proper supervisions of the examinees and calibration and maintenance of apparatus, a ratio that makes widespread testing very resource intensive, and thus lowers the utility trade-off with test validity.

*Note. Based on edited excerpts from Ackerman and Cianciolo (1999, pp. 232-233) who conclude the application of digital technologies to psychomotor testing eliminates all of these barriers.

Procedural Skills. Procedural skills are also an excellent example of the need for and potential of digital technologies to provide more precise human performance measures. Procedural or rule-based behavior is fundamental to most human performance, particularly in familiar situations. And in highly structured organizations and job environments, such as the military, the need for precise measures of procedural performance might seem eminently clear. Data on procedural performance should be needed to assess and validate the selection and promotion of personnel, and their training, including the acquisition, retention, and transfer of procedural skills essential to job performance. Computer-mediated work not only enables procedural data tracking, it also reinforces and often requires procedural compliance.

Precise job performance data at the procedural level, however, is rarely collected or available. This is particularly true of procedural job performance data in many work environments, including the military. "Unfortunately, hands-on performance is seldom measured or publicly available…" was a conclusion of Army researchers in their efforts to validate enlistment test batteries against actual job performance (Sands, Waters, & McBride, 1999, p. 277). With the exception of special studies, they surmise "hands-on performance is nearly invisible to external decision makers…." Moreover, direct measures of procedural performance are rarely collected, particularly under realistic job conditions. For example, recall that the training evaluation formats used by Observer/Controllers often require only "Go" or "No Go" ratings at task and subtask levels.

Procedural behaviors, however, can be and often are readily tracked by digital technologies, as discussed under Digital Instrumentation. Moreover, as the range and importance of work performed on computers expands, the performance logged equates to the procedures required to perform jobs. Precise and comprehensive compilations of computer-based procedural data, therefore, might support a wide range of training and evaluation goals ranging from linking process and product performance with overall outcomes, to providing training analysis and performance feedback. For example, Kontogiannis (1999) examined how digital tracking data can provide timely feedback on emergency operating procedures *during* an emergency.

The analysis by Kontogiannis (1999) began by assessing procedural limitations in handling nuclear power plant emergencies due to paper copy media. The analog limitations identified included: presenting complex instructions, handling cross-references, tracing suspended or incomplete steps, and monitoring user progress. This work examined how on-line procedures were currently used at more than a dozen nuclear power facilities. These facilities use a digital tracking technique called process linking that digitally tracks workers' procedural process and links that process directly to the computer-monitored work environment, including emergency conditions. The analysis concluded that digital process linking provided concurrent feedback to plant operators that improved their ability to handle emergency on-line procedures.

The work by Kontogiannis also provides a research framework for understanding and categorizing the potential uses of procedural data for training and evaluation. Behavioral categories in this framework for monitoring task progress and error management, for example, include: monitoring parallel tasks, logging and notifying incomplete and interrupted tasks, tracing task completion, and previewing tasks that must still be performed. Procedural frameworks such as the one proposed by Kontogiannis should apply to many different work settings and provide a coherent and more uniform structure for evaluating and improving computer-based procedural performance.

Cognitive Skills. Cognitive skills are a particularly interesting example of the need for and potential of digital technologies to provide more precise human performance measures. Surprisingly, perhaps, many important aspects of cognitive performance can be directly measured by digital technologies. Striking examples of computer-based measures of cognitive performance measures range from school and job testing to neuroimaging. More germane examples, with respect to command and control cognitions, include the ability to track and critique how humans access, filter, compare, integrate, coordinate and apply information.

In the area of testing, for example, digital technologies alter both what is appropriate for learning and testing, and what is in fact "testable." Educational theorists and practitioners often extol the need to learn, and assess learning, in more authentic situations. Authentic situations are purportedly conducive, if not inductive, to constructing and/or reconstructing a more useful knowledge base. However, the contextual and informational affordances provided by digital technologies in the creation of more authentic situations, may radically change learning and performance requirements. A telling assessment of how digital technology redefines the appropriate content for cognitive learning and testing is:

The interactions of hardware, software, and interface innovations can provide for some long-lasting and solid changes in educational environments. Changes in scale of information, in the tools of information manipulation, in the cost of widely distributed databases, and in the "appearance" of information, promise revolutions in the human processing of information; education can focus on human thought and thoughtfulness in a range of media, and defocus from the memory of textual facts as the current very new potentials are realized. (Hooper, 1988, p. 322)

Digital technologies also expand the set of methods and measures for assessing cognitive processes and skills. In testing situations, for example, computers are being used to measure test-taker's mouse movements, the elapsed time between item presentation and test-taker's response, and the underlying cognitive processes used to answer items (Burroughs et al., 1999). Computer-measured processes may include what information was requested before the test-taker responded, the order in which the information was provided and examined, and the amount of time spent consulting that information. Test developers are no longer limited to multiple-choice questions, an assessment format that has long been a source of concern for the measurement community, Burroughs et al. states. Similar observations were made by Sands et al. (1999) who predicted that computer testing is poised for a broad technology transfer to the entire spectrum of testing, particularly cognitive testing. In fact, they concluded that the conquest of computerized testing is "inevitable."

Two examples of how computers are being used to assess cognitive performance in applied work settings are presented here. The first was an evaluation of workers performing the task of identifying antibodies to determine safe blood donations for patients (Guerlain et al., 1999). Workers used a direct manipulation interface that provided an unobtrusive form for monitoring their problem solving process, and detecting potential errors in their identification procedures. By applying a cognitive procedural model, the computer was able to critique decisions and provide feedback on errors of omission and commission, and ensure an accurate identification. Their evaluation effectively demonstrated the utility of computer-based monitoring of cognitive processes to improve performance on a difficult and critical task.

Perhaps, more importantly this effort resulted in an explicit set of technology design principles for monitoring cognitive processes, and interactively linking worker and computer skills in support of real-time job performance. These design principles include: use direct manipulation to provide an unobtrusive form of communication; provide perceptual and memory aids to encourage the user to use the system's direct manipulation capabilities; base the critiquing strategy on an error model for the domain; and, abstractly represent the computer's knowledge for the user to establish a common frame of reference. A good summary of how they designed a system that tracks and critiques cognitive performance is:

With an interactive critic, the system monitors all of the user's data collection and interpretation activities, and provides immediate feedback when it sees a potential problem. To be effective, the critiquing system must be designed so that it has access to data about the user's intermediate cognitive processes while he or she is solving the problem. (Guerlain et al., 1999, p. 75)

A final example of cognitive tracking directly related to command and control performance is recent work by Ehret et al. (2000). Their work addressed the information processing performed by submarine approach officers attempting to localize an enemy submarine hiding in deep water. The authors stressed their study targeted cognitive processes, not performance outcomes; not how well officers do, but how they do well. More precisely, their goal was to derive a step-by-step record of the information processing invoked during a critical stage of the officers' job, situation assessment. To do so the authors used what they called a scaled world model. A scaled world serves as a middle ground between the situational complexity inherent in field research that resists definite conclusions, and the situational paucity of laboratory research that defies useful conclusions. It preserves key functional relationships based on questions of interest to evaluators or trainers, while paring away other functions that might confound answering those questions.

Initially, problems in analyzing cognitive performance with a high-fidelity simulation of the ocean environment were identified by observing officers' interactions with the simulation displays as they localized enemy submarines. Problems restraining a more precise analysis by the researchers included: nonunique information sources where the same information could be obtained from multiple displays; amount of information per display where copious information on the displays encouraged information browsing that was hard to track; and, receipt of unrequested information from intermediate displays viewed while navigating from one display to another.

To overcome such problems, the researchers designed and developed "Ned," a scaled-world version of a high-fidelity ocean environment. The design for "Ned" modified functionality to provide a more precise tracking of the query and receipt of information while reducing receipt of unrequested information. For example, "Ned" reduced nonunique information by making information fields and interface actions unique to a given display, and reduced receipt of unrequested information by providing an always-visible palette menu that enabled direct navigation between the 10 displays available, without intermediate menus and displays. The instrumentation of "Ned" also addressed the need for more precise data on information processing performance. For example, instrumentation logged and time-stamped each officer action on the interface, the content of the information fields viewed, and the duration in milliseconds the information was visible.

An initial analysis of cognitive performance tracked the information processing of 36 officers and instructors with 23 years in the Navy, with 6.4 of those years at sea. During an approximately 18 minute scenario, each participant solved a series of problems related to detecting hostile submarines. As Ehret et al. note, "The problem is always the same: What is the state of the world now?" (p. 14). The design and instrumentation of "Ned" allowed researchers to automatically encode 350-450 information actions by an officer, per scenario. In addition, the log file automatically segmented these operators into officer goals and subgoals with 95 percent accuracy. For a prototype effort, the ability of "Ned" to automatically track cognitive performance and provide a step-by-step relation of that performance to problem solving goals and goals appears impressive. One might question how valid and useful an automated analysis of cognitive skills might be?

Concerning validity, many of the highly experienced participants rated the scaled world of "Ned" as both realistic and engaging, and improvements are planned. Scaled worlds by design restrict external realism for the sake of internal validity. However, the authors remind us that a preoccupation with mundane realism may restrict generalizability in an ever-changing world. Moreover, empirical comparisons between scaled and real world performance are supported. For example, data collected on where officers believed the enemy submarines to be located can be played-back and compared with actual locations. In addition, computational models of the operators used by the participant experts to solve scenario problems are being developed. In what the authors call a Turing test, independent experts will compare the log files generated by senior officers, junior officers, and the computational model to determine how well "Ned" succeeded in measuring cognitive expertise.

The potential utility of cognitive tracking approaches, such as "Ned," for training and evaluation applications is of particular interest for this report. For example, trainees might readily benefit from codification, and particularly simulation, of expert problem solving across diverse and authentic job scenarios. In turn, evaluators would have a precise, step-by-step protocol for assessing and comparing cognitive performance, at expert, intermediate, and novice levels. And as demonstrated by Guerlain et al. (1999), a digital model of experts' cognitive performance would enable interactive critiquing and real-time feedback on both trainee and worker performance. At a more basic level, cognitive tracking methods should contribute to the understanding and development of expertise, including command and control expertise, by empirically examining and advancing cognitive theory.

Collaborative Skills. Collaborative skills are briefly examined as a final example of the need for and potential of digital technologies to provide more precise human performance measures. The network centric nature of digital technologies has resulted in an unprecedented extension of collaborative performance. Interactive group sizes now range from organizational teams and structures to literally global information networks and global observatories on the World Wide Web. For example, during the opening ceremony for the '98 Winter Olympics in Nagano, chorus groups on five separate continents sang in "real-time" unison the first global song. As digital technologies advance, group sizes are expanding and the pace of interaction is accelerating. Analog and manual measurement methods are totally incapable of collecting collaborative data given the volume of information associated with networked groups, and incapable of updating collaborative databases given the pace of digital interaction and information exchange rates. Strategies and methods for digitally tracking and assimilating collaborative performance, however, are rapidly emerging. Two examples of how digital measurement methods help assess collaborative performance in applied learning settings are provided next.

The first example is an effort to scale-up the assessment of computer-supported collaborative learning for engineering students (Guizdal & Turns, 2000). While these authors report positive learning outcomes from technology-based student collaboration, such as Internet discussion groups, they were concerned with the need to more directly monitor the process of collaboration and ensure responsive feedback was provided. They describe a typical scenario for an introductory engineering class of 50 students contributing an average of three collaborative "notes" per week, totaling more than 1500 notes over the term not including teacher and moderator notes. Careful analysis of each student dialogue or note is important at the individual level. However, understanding how the class overall is engaging in the collaborative process requires an examination across notes and contributors. The scope and scale of their evaluation challenge was summarized thus: "…the focus of the evaluation scales from microanalysis of individual contributions to macroanalysis of activity in groups of discussions with thousands of notes and tens to hundreds of participants" (Guzidal & Turns, 2000, p.229).

To scale the evaluation to the level of interactive dynamics, the authors relied on the integrative nature of digital technologies to identify "threads" of communication. Threads are a tracing feature that strings together a collection of notes created in response to some initial note, and threads are common to many web-based news and discussion groups. Thread-based measures used by Guizdal and Turns to characterize class discussion included: number of threads, average thread length, number of notes per thread, number and percentage of notes written per contributor, and number of notes written in response to others' notes. Their analysis of this data resulted in a conclusion that substantial and substantive collaboration across many students occurs in most of the engineering environments examined. This example indicates how digital technologies provide a multilevel approach to learning and evaluating at micro individual and macro group levels.

The second example of how digital measurement methods help assess collaborative performance concerns distance learning using groupware (Helms, Neale, Isenhour & Carroll, 2000). This effort used a server-based approach to logging data to achieve higher-level capture and multi-level abstracting of collaborative activities. A server-based approach in essence is independent of the application being used at the client level, and has very limited access to the relatively low-level individual user/client interactions such as keystrokes and mouse movements available with fully instrumented systems. In contrast, server-based logging captures multi-user/collaborator actions at the level of changes to shared data, such as a groupware product.

The authors stress that server-based logs map more directly to meaningful human behavior than client-based logs. They devised a three-tiered model to categorize the collaborative behavior tracked into more meaningful categories: user moves (such as opening software tools, browsing, and initiating conversation); artifacts generated (changes to shared products such as drawings, text documents, and databases); and, human-to-human communications (such as chat messages, e-mail, and video conferencing). By coding server-level events into these behavioral categories the authors report they were able to build an integrated chronicle of system events and related collaborative behaviors the authors called "integrated activity scripts." The authors concluded that the completeness of the information and identifiers contained in these scripts allowed them to generate a variety of meaningful quantitative measures of collaboration, as well as a means for organizing script content and related occurrences for qualitative analysis.

Overall, the potential for increased precision and scope in measurement methods by applying digital technology to track human behavior includes psychomotor, procedural cognitive, and collaborative performance. This report stresses that the precision and scope of such data provides unprecedented opportunity for understanding and improving human performance.

However, the profusion of digital performance data even currently available is potentially overwhelming and counter productive. Fortunately, methods are rapidly emerging to overcome these problems. Answers to these problems include better methods for integrating, mining and visualizing vast databases. These emerging answers are considered later under "Examples of Digital Measurement Methods." A common component to all of these answers is that digital technologies can help resolve evaluation problems, including the problems they create.

More Meaningful Evaluation

Successful performance and meaningful evaluation depends on the situation or context. The performance setting is a major determinant of what behaviors are performed when, where, and why, as well as the evaluative frame of reference required to assess how well behaviors were performed. Fortunately, researchers and practitioners are increasingly aware how important the situation is to understanding human performance. This awareness is underscored by a recent review and analysis of performance reported by Kirlik and Bisantz (1999). Notably, their analysis focused on areas of cognitive performance central to command and control: planning, problem solving, and decision making.

Empirical findings increasingly indicate that performance is rarely based on a detailed and deliberated plan of action, but rather on a progressive cycle of situation perception and action (Kirlik & Bisantz, 1999). Research in more contrived and low-fidelity situations, including traditional laboratory efforts, has failed to adequately predict behavior in more natural, higher-fidelity, situations. In contrast, research conducted in more natural situations, such as job settings, demonstrates that situated action is more ad hoc. Situated actions are responsive to the momentary changes that occur in dynamic situations.

Accordingly, a major premise of this report is that digital media *situate* performers in the performance context. A fundamental attribute of media, an extension of the self, is elevated by digital technologies to new levels. These levels reflect what DiSessa (1986) referred to as the digital "trick" of turning abstractions into experience, and what is more commonly referred to by terms such as "virtual" reality. Instances of virtual realities are increasing and include computer-mediated learning contexts (DiSessa, 1986; Guzidal & Turns, 2000) and training contexts, particularly virtual simulation (Fletcher, 1994).

Digital technologies can do more than provide compelling representations that induce experience in realistic performance settings. Digital technologies routinely provide microscopic and macroscopic levels of reality beyond human sensory thresholds. Common examples of these levels include medical and satellite imagery. Such representations provide more complete and meaningful performance settings that augment perception as well as significantly expand, and often improve, the range of performance possible.

Digital representations also help us overcome experiential misperceptions. For example, the work of DiSessa (1986) focuses on how real-life experiences are confounded with invisible or imperceptible factors that often distort or contradict meaningful understanding of the actual situation. The world of sensory experience is, in fact, misleading with respect to Newton's Laws about invisible forces such as gravity and friction. DiSessa's work, therefore, uses what he calls "immersive" settings that allow learners to feel as if they are "inside" objects moving through a world in which gravity and frictional forces are set to zero, in order to overcome misperceptions and misconceptions based on a lifetime of normal experience.

A final consideration is that the line between virtual and real world performance settings is becoming increasingly blurred. As the computer interface becomes an increasingly accurate and meaningful representation of reality as well as a viable link to reality, it is becoming the users' primary interface to the performance setting. The expanding array of microworld settings ranges from electronic commerce to digital command and control systems. In sum, computer-mediated environments for training/work are becoming indistinguishable, if not identical, and literally situate performance in the work settings in which it occurs.

This report contends that the potential of digital media to enable more meaningful evaluation resides primarily in its ability to relate performance to context. The preceding discussion has stressed that the situation is essential to understanding behavior, and that the situational representations provided by digital technologies are increasingly capable of and indispensable to communicating meaningful performance contexts. The report's prior review of automated measures of human performance provided numerous examples of the ability of digital technologies to create, maintain, and sustain performance contexts ranging from psychomotor skills to cognitive and collaborative efforts.

This report's definition of C^4I instrumentation underscores performance context: instrumentation is defined as a log of all soldier-computer interactions *correlated* with the battlefield situation in which they occur. The conjunctive nature of digital technologies, sometimes described as network centric systems, readily supports the requirement to relate performance to context. Initially, this entails the ability to log or track all soldier-computer interactions at the individual level. In addition, given the collective and collaborative conditions of military operations, this ability extends data logging across C^4I systems to include any and all C^4I equipped combatants and supporters. Finally, the digital nature of military simulations and particularly soldier-in-the-loop virtual simulations extends data logging and data synchronization to include weapon and sensor system performance as well as simulated battlefield conditions such as mission, terrain and enemy information.

A corollary premise is that digital media also situate *evaluation* in the performance context. The ability of digital evaluation technologies to relate performance to context provides evaluators, as well as performers, the situational conditions and cues that determine performance. This situational data, including future situation information in the form of mission objectives and desired end states, provides frame of reference data essential to meaningful evaluation. In addition, digital technologies and particularly simulations afford trainers and evaluators the ability to control contextual variables and customize scenarios to more precisely meet training objectives and evaluation issues. Fletcher strongly advocates the potential of instrumented digital work settings to provide more meaningful evaluation: "... instrumenting the electronic battlefield... provides the foundation for a measurement system that should substantially advance our assessments of crews, teams and units in both military and nonmilitary settings" (Fletcher, 1994, p. 268). Overall, digital technology affords the situational data needed to understand the purpose of performance, and the adequacy of performance relative to purpose.

Less Burdened Measurement Methods

A primary challenge is how to reduce the high labor and time costs invariably associated with evaluation of conventional command and control performance, previously examined under Burdened Measurement Methods. This challenge is magnified by the opposing need to improve the precision and scope of performance data collected in order to reverse a pattern of sparse assessment and uneven quality in command and control evaluation. In addition, it was proposed that analog media contribute heavily to the burden associated with performance, and particularly evaluation, of command and control.

This section examines how digital evaluation technologies could result in less burdened command and control measurement methods. This examination focuses on the potential of digital technology to impact measurement, particularly observer workload. As preface, we recall McLuhan's assessment about the potential impact of media on observation: "The older training of observation has become quite irrelevant in this new time, because it is based on psychological responses and concepts conditioned by the former technology..." (McLuhan, Fiore, and Agel, 1967, p. 8).

The previously considered analytic efforts by Brown et al. (1998) documented the high workloads experienced by observers and analysts at the Army's premiere training centers. Moreover, their analysis also included an examination of how digital technologies, and particularly evaluation and instrumentation technologies, might reduce observer and analyst workload. The rationale for this extended analysis into workload reduction was based on their stated assumption that unless training centers are equipped with digital evaluation technologies they simply cannot meet the training feedback requirements introduced by force modernization initiatives. The modernization initiatives analyzed focused on equipping units with C^4I systems, but also included other advances in a wide range of weapon and sensor system modernization efforts. Their focus on the potential for increased evaluator workload was underscored by their assessment that the information generated by C^4I systems would simply "overwhelm" observers and analysts (Brown et al., 1998).

To support their workload reduction analysis, Brown et al. postulated 13 strategies designed to improve the Army's current live simulation and instrumentation systems. Overall, these strategies stressed the ability of digital technologies to automate data collection, analysis, and presentation efforts. Selected strategies, for example, were: Automate C^4I Data Collection and Control, Automate Tracking of Player Activities and Expended Resources, and Automate AAR Preparations. Each strategy and its specific impact on training feedback requirements are documented in their report (Brown et al., 1998).

The results of their analysis strongly endorsed the potential of digital technologies to reduce evaluators' workload. Overall, they concluded that of the 380 tasks performed by observers and analysts that they identified and analyzed, implementation of all 13 strategies would result in full to partial workload reduction for 368 tasks. Although the workload associated with 97 percent of these evaluator tasks might be reduced, that does not equate to a 97 percent reduction in overall workload given that some tasks would be only partially automated. However, their analysis provides a compelling and relatively comprehensive assessment of how digital instrumentation and simulation technologies might result in less burdened measurement methods.

In addition, it is important to note the workload reductions estimated by Brown et al. (1998) are based on a limited notion of instrumentation, and particularly C^4I instrumentation. Their more limited notion of C^4I instrumentation might be characterized by a reliance on *inter*-system, or server-based, information exchanges. Based on this level of instrumentation, they assume that evaluators will know what information was available to and/or received by performers equipped with C^4I systems. With only inter-system data, however, evaluators will not know if the performers examined that information or what they did with it. They limited their analysis to this level of instrumentation because of perceived technical limitations. Current technologies, particularly the limited bandwidth available for wireless communications including C^4I systems in a live environment, may not support the collection of intra-system data.

A more comprehensive concept of C^4I instrumentation that includes *intra*-system data, however, might reduce evaluator workload beyond that projected by Brown et al. (1998). For example, one of most important and onerous evaluator tasks, particularly for conventional command and control evaluations, is the manual reproduction and comparison of users' battlefield representations. Consider how C^4I display data might impact evaluation workload. By design, our proposed definition of C^4I instrumentation includes the collection of intra-system data. A log of all soldier-computer interactions includes a record of all soldier and system inputs and outputs at each user's workstation or C^4I system. For example, an intra-system log should include a record of display outputs, or the data required to reconstruct the information depicted on a user's C^4I display at any or all times during a recorded exercise.

In contrast, digital evaluation technologies should be able to automatically collect and compare the battlefield representations depicted on users' instrumented C^4I systems. A more detailed examination of this capability, referred to in this report as automated pictorial comparison, is provided in the next section titled Examples of Digital Measurement Methods. Here, we simply stress that instrumentation that captures intra-system data greatly expands the precision and scope of measures important to command and control, such as a comparison of users' situational representations. And although evaluation workload and efficiency are the immediate focus, the potential impact of intra-system data on evaluation effectiveness is always a concern.

Intra-system data provide evaluators a much more accurate and discriminating account of information access and utilization than inter-system data alone. It greatly overcomes many of the data shortcomings noted by Brown et al. (1998) such as not knowing if the performers examined the information they received or what they did with it. For example, work by Brown, Metzler, Riede and Wonsewitz (1996) provides excellent examples of pictorial comparisons that target what they refer to as "ground truth" (the actual situation) versus "perceived truth" (the situation as perceived). Such comparisons are speculative when based only on inter-system data, however. Inter-system data disclose only the information that *should* have been available, based on a log of the messages received by a user's C^4I system. However, these messages might not have been correctly received, might not have been opened and displayed, or might not be directly visible in the map area shown on a user's C^4I system at the time of comparison. Whereas, comparisons based on intra-system can disclose precisely what information was *actually* available, and even visible, on each user's C^4I display at any moment during a simulated battlefield exercise.

An assessment of the technical feasibility of logging intra-system data with fielded C^4I systems is admittedly beyond the expertise of this report's authors. In the information age, however, what seems difficult to impossible today, is often commonplace tomorrow. Moreover, much of the research previously reviewed under the Instrumentation section was based on intra-system data collected in actual computer-mediated work environments (e.g., Guerlain, 1999; Ehret et al., 2000; Kontogiannis, 1999).

Moreover, the Mounted Warfare Test Bed (MWTB) at Fort Knox, Kentucky, has developed and tested fully instrumented C^4I systems integrated with virtual simulation technologies. Research conducted in this test bed has repeatedly demonstrated the potential of digital evaluation technologies for capturing and correlating digital C^4I and simulation data (Leibrecht, Meade, Schmidt, Doherty, & Lickteig, 1994; Throne, Deatz, Holden, Campbell, Sterling, & Lickteig, 1999). Examples of this work and related digital measurement methods that should reduce evaluator workload and improve evaluation effectiveness are provided in the following section. In conclusion, what the Army has demonstrated as technically feasible in virtual training and evaluation test beds today, may be feasible in live environments tomorrow. Where we can improve training and evaluation today, we should not wait till tomorrow.

Examples of Digital Measurement Methods

This section of the report provides some selected examples of digital measurement methods and automated measures. These examples are presented to help examine and illustrate the potential of applying digital technologies to evaluation, and particularly to evaluations of command and control performance. Some of these examples are based on research and development efforts conducted by Army Research Institute working directly with the Mounted Maneuver Battle Lab (MMBL) at Fort Knox, Kentucky, and primarily the MWTB. The other examples are drawn from related military research, including ground and air operational settings. The examples are organized under three key issues that address both the potential of digital evaluation technologies and the requirement for additional research and development in order to realize that potential: data integration, data mining, and data visualization.

Data Integration

The inherent ability of digital technology to integrate data and information is central to the Army's ongoing efforts to improve command and control performance and evaluation. The Army's definition of digitization underscores the need for this integrative ability: the application of technology to acquire, exchange, and employ timely information horizontally and vertically integrated to create a common picture of the battlefield from soldier to commander (U.S. Department of the Army, 1998b, p. 5). However, the objectives of Army digitization to mine data to discover information and to visualize data to "see" the battlefield, depend heavily on the emerging ability of digital technology to collect and integrate the data and information requirements of future forces.

With the advent of digital systems, the world is awash with data and digital information processing technologies. For instance, a recent analysis by the University of California, Berkley concludes that the amount of unique data generated yearly in the new digital age is around 1.5 exabytes (1,500,000,000,000,000,000 bytes) (Schorow, 2000). Individuals create most of this data in the form of documents, photographs and home videos. Satellites stream huge volumes of data about such things as earth conditions, including oceans, atmosphere, and topography. In the area of electronic commerce, bar codes and scanners constantly record and consolidate information about what is bought when, where, and by whom. And commercial, academic, and government organizations are actually creating, versus just talking about, "paperless" environments.

This relatively recent torrent of digital data not only overwhelms human information processing and integration abilities, but also for the most part resides beyond human observation. Computers increasingly operate at the front-line of data collection, and exercise growing control over access to the data they collect, notes Bailey (1996). He examines how this trend of computer controlled data collection, processing, and presentation affects areas such as finance, space exploration, and military operations. Financial data, for instance, are routinely handled by computers first, and humans later, if at all. In a fighter airplane, sensor and computer circuits acquire and detect enemy aircraft long before the pilot, and often have completed identification estimates by the time information on the existence of the enemy plane is provided to the pilot.

Fortunately, digital technologies are increasingly capable of collecting and integrating data. Our information processing today is estimated to be a trillion times greater than at the dawn of civilization, with almost all of that increase being electronic (Bailey, 1996).

Computers are essential, in part, because of the sheer enormity of the data available and the growing need for integrative processes, including correlation, association, and fusion. As Parsaye and Chignell (1993) stress, there will never be enough scientists and evaluators to explore the boundless world of data being assembled by digital technologies. Computers are also essential, however, because the nature of the problems we face and the information we hope to discover are changing. As Bailey (1996) notes, the cutting edge of science has shifted from nonadaptive to adaptive, evolving domains such as immunology, marketing, and intelligence. In such areas, humans' "...rational thought of place and pace... (Bailey, 1996, p. 214)" does not penetrate very deeply into the perpetually changing world of pattern. Computers are uniquely suited to explore the actual versus our abstracted patterns of the real world, however, and discover what is literally beyond thought.

Realization of an integrated military database is crucial to the Army's digitization effort designed to meet the information requirements of future forces. Most practitioners and designers understand that the common picture objective does not equate to an identical picture for all users, if any. Rather, the term "common" underscores the requirement that the display provided to all users is based on the same data—from the commander in chief to the soldier in a fighting vehicle (Boller, 2000). Within the Army, the obstacles to achieving this level of data commonality are severe and far from overcome. One obstacle is the requirement for unique and predefined organizational structures and identifiers for every potential user. Boller describes Army efforts that defined a default organizational chart that identifies and links all relationships within an organization. To apply this default organization, Boller states the Army must now adapt it for 4,900+ modified tables of organization and equipment and tables of distribution and allowance for the Active and Reserve Components.

Another major obstacle is that the Army's currently fielded set of C^4I systems do not communicate adequately with one another and, therefore, do not contribute to an integrated database. The Army's transition to a truly integrated digital system, currently referred to as the objective Army Battle Command System, is a stepped process that is eventually expected to result in the Army's Common Database. A related, more formidable challenge is the growing requirement for joint service operations and interoperability. An integrated database across services is essential to providing a common picture capability to future forces.

> *The key to the science of seeing and interoperability is the Joint Common Database, a fully integrated, distributed database that all automated command and control systems use to share information. It is not a "big database in the sky," but tailored to each organization in content, size, area of coverage and overlay features.* (Boller, 2000, p. 38)

Fortunately, the ability of digital technologies to collect, integrate, and update common databases is increasingly evident. Digital technologies routinely fuse data into what Parsaye and Chignell (1993) describe as "intelligent databases" and these integrated databases currently support a wide range of applied settings. Logistics, for example, is a critical requirement for achieving a common military database. Parsaye and Chignell describe how a grocery logistics applied a digitally integrated database as early as 1990:

> *...10,000 route salespeople in one company used handheld computers to record daily sales information about 200 grocery products in 400,000 stores. The stored information was then transmitted nightly to a central computer which in turn returned pricing and product promotions to the hand-held computers. The resulting data was then combined with external data about the sales of competitive brands on a weekly basis and this information was summarized and provided to executives through an executive information system.* (Parsaye & Chignell, 1993, p. 327)

More current examples of logistic data integration include the ability of commercial shipping businesses to track package locations in near-real time and provide that information on-line to both corporate personnel and individual customers.

The Army's efforts to develop more integrated digital databases include a wide range of research and development efforts. Currently fielded C^4I systems are not fully integrated and do not as yet result in a common database. However, Army digitization efforts include research and development with more advanced or "objective" C^4I systems conducted in more robust and integrated digital data environments.

The Army's MWTB provides a powerful example of how digital technologies such as virtual simulation and instrumented C^4I systems can provide an integrated digital data environment that supports training and evaluation. Research conducted in this test bed has repeatedly demonstrated the potential of digital evaluation technologies for capturing and correlating soldier-machine performance data collected during simulated operational conditions (Leibrecht et al., 1994; Throne, Holden & Lickteig, 2000). The digital infrastructures developed in the MWTB automatically collect and integrate soldier-in-the-loop simulation data and soldier-computer interaction data, as depicted in Figure 3.

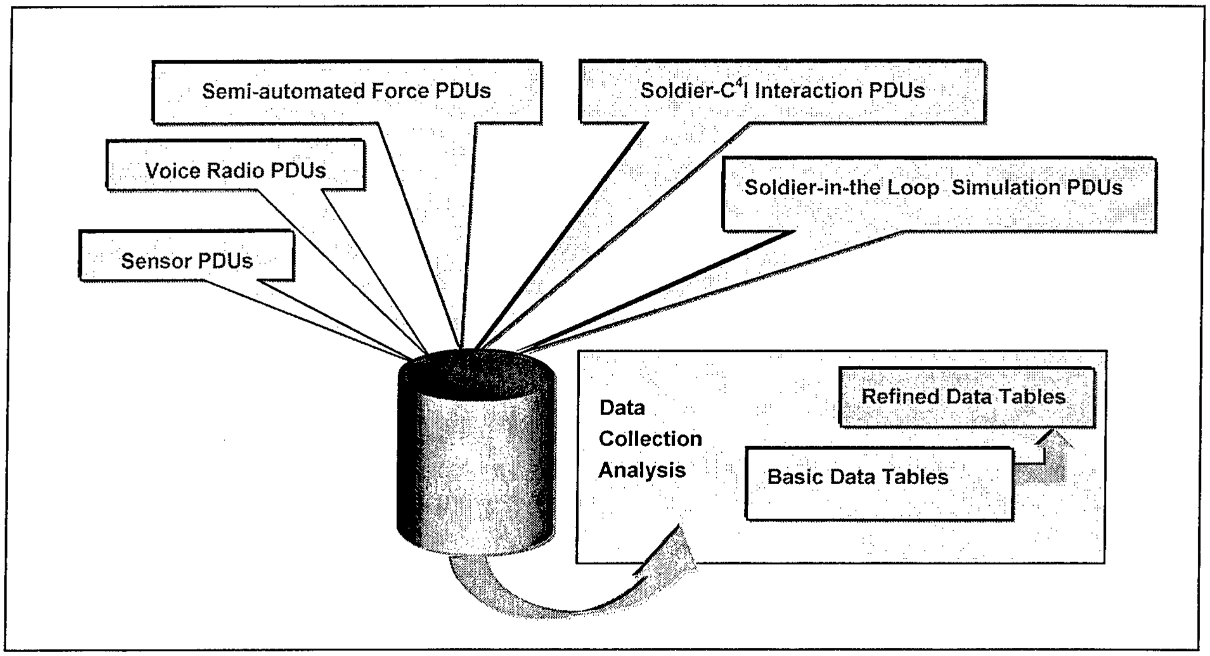


Figure 3. An example of integrated digital data environment in the Mounted Warfare Test Bed.

The purpose of this section is to present selected examples of how digital data integration technology has and might be used to support command and control performance and evaluation. Three examples of digital data integration are provided below. The first two examples are directly related to command and control training and evaluation efforts. These examples illustrate how the MWTB's integrated data digital environment supports the development and application of automated measures of command and control performance. The third example is based on a digitally integrated evaluation test and evaluation system developed by the Air Force to provide evaluators an integrated multimedia data analysis capability. A summary discussion of these data integration examples is provided at the end of this section.

Combat Vehicle Command and Control. The first example of digital data integration and measurement methods is based on a program of research titled Combat Vehicle Command and Control (CVCC). The CVCC program evaluated simulation-based platoon, company, and battalion units equipped with a prototype objective C^{4}I system compared to baseline units equipped with conventional or analog command and control components. The virtual simulation environment for these evaluations was a form of distributed interactive simulation called Simulation Networking (SIMNET) that was first installed in MWTB (Alluisi, 1991).

From an evaluation perspective, a key component of the MWTB infrastructure was the ability to record and analyze a wide range of evaluation data the test bed's data collection analysis (DCA) system, as indicated in Figure 3. From a command and control evaluation perspective, another key component was instrumentation of CVCC's prototype C^4I systems. Soldier-computer interactions with these C^4I systems were also recorded by the DCA and automatically correlated via time stamping with all other exercise data, including simulation data, in the form of protocol data units (PDUs). For a more complete description of the CVCC program, including automated measures and methods, see Leibrecht et al. (1994).

The experimental designs and measurement methods of CVCC targeted evaluations of training, soldier-machine interface, and operational effectiveness. The overall set of CVCC measures were organized under four battlefield operating systems (BOS): Maneuver, Fire Support, Intelligence, and Command and Control. For example, a sample of the CVCC measures used to assess command and control performance is provided in Table 4.

These measures addressed the reception and transmission of mission, enemy, and friendly information, as well as information management. For the experimental units equipped with prototype C^4I systems, most of the CVCC command and control measures were automated measures due to C^4I instrumentation. For the baseline condition, all command and control measures were obtained manually. For both conditions, measures based exclusively on voice versus digital reports such as requests to clarify communications were obtained manually from voice data recordings captured by the DCA, as indicated in Table 4.

Table 4

Sample of Command and Control Automated Performance Measures from Combat Vehicle Command and Control Battalion Evaluation

Command and Control Function	Sample of Command and Control Automated Performance Measures from Combat Vehicle Command and Control Battalion Evaluation
Receive and Transmit Mission Information	Mean time for transmission of FRAGO across echelons (battalion, company, platoon). Duration of requests by company and platoon members to clarify. Consistency of FRAGO received. Number of requests by company and platoon members to clarify. FRAGO/overlay*.
Receive and Transmit Enemy Information	Time to transmit Intelligence Report (INTEL) across echelons. Consistency of INTEL received across echelons. Number of requests to clarify INTELs by company and platoon members*.

Table Continued

Table 4 (Continued)

Receive and Transmit Friendly Information	Mean time to transmit Situation Report (SITREP) across echelons. Mean duration of voice transmissions within and between echelons. Deviation of BLUFOR locations reported in SITREP from actual location. Delay between observed phase line or line of departure crossing and reported crossing. Delay between observed battle position arrival and reporting set at battle position. Elapsed time from request for fuel and/or ammunition report across echelons. Number of voice transmissions within and between echelons.
Manage Means Of Communicating Information	Average length of voice radio transmissions by echelon. Total number of voice radio transmissions by echelon. Total time on voice radio network by echelon. Number of named voice reports by echelon*.

* Indicates manual versus automated measures primarily for voice data recordings.

Of special note, the MWTB's integrated data environment helped CVCC relate command and control performance to unit MOPs and MOEs. For example, Table 5 provides a sample of the CVCC automated measures used to assess Maneuver performance. Unit MOPs in Table 5 that are closely related to command and control performance include measures under the Maneuver categories of: Move On Surface, Navigate, and Process Direct Fire Targets. Unit MOEs in Table 5 are primarily engagement-based measures identified under Engage Direct Fire Targets. Nearly all of the Table 5 measures are based on simulation data that were automatically collected and compiled by the DCA for both digital and baseline units during CVCC. Although beyond the present scope, recall that automated BOS measures for Fire Support and Intelligence were also developed to provide a more complete and integrated CVCC database for command and control evaluation.

Table 5

Sample of Maneuver Performance Automated Measures from Combat Vehicle Command and Control Battalion Evaluation

Maneuver Function	Sample of Maneuver Performance Automated Measures from Combat Vehicle Command and Control Battalion Evaluation
Move On Surface	Distance between BLUFOR and OPFOR center of mass. Time to reach line of departure. Average number of OPFOR vehicles to which each BLUFOR vehicle exposed. Range to OPFOR at displacement. Time for companies to reach objectives.

Table Continued

Table 5 (Continued)

Process Direct Fire Targets	Mean time to acquire targets. Mean time between lases to different targets. Mean time from first lase to first fire. Maximum lase range. Number of fratricide hits by manned vehicles. Number of fratricide kills by manned vehicles.
Navigate	Mean distance traveled. Mean fuel used. Mean time out of sector/axis. Mean time misoriented. Time to complete mission/mission segment.
Engage Direct Fire Targets	Percent of OPFOR killed by end of stage. Percent of BLUFOR killed by end of stage. Losses/kill ratio. Mean target hit range. Mean target kill range. Percent OPFOR vehicles killed by all manned vehicles. Hits/round ratio, manned vehicles. Kills/hit ratio, manned vehicles. Kills/round ratio, manned vehicles. Number of manned vehicles killed. Number of rounds fired by manned vehicles. Number of OPFOR vehicles killed relative to designated Phase Lines.

A final category of automated measures assessed how soldiers used their prototype C^4I systems. A sample of the CVCC measures used to assess soldier-computer interactions is provided in Table 6. These measures included data on map scale and scroll interactions, and reports received and relayed. Again, these measures leveraged the fact the C^4I systems used during CVCC were designed and developed as instrumented systems. And all of these soldier-computer interaction measures were automated measures that applied only to soldiers equipped with prototype C^4I systems. Of special note, C^4I instrumentation supported the collection of both inter-system and intra-system data, and soldier-computer interaction measures based on each of these data types are identified in Table 6.

Battle Command Reengineering. The second example of digital data integration and measurement methods focused on command and control performance is based on a research program titled Battle Command Reengineering (BCR). The Army Concept Experimentation Program (CEP) included a series of BCRs (I-IV) conducted in the MWTB from 1997-2000 that examined the effects of advanced C^4I systems on battle command reengineering. The BCRs seek to create future battlefield conditions that might exist in 2012 and beyond. These conditions include: completely integrated C^4I systems, both vertically and horizontally, that perform routine information collection and dissemination tasks; target and intelligence reporting capabilities seamlessly integrated into those information systems; lighter, more mobile weapons systems with increased range and lethality; and robotic systems to support reconnaissance and supply.

Table 6

Sample of Soldier-Computer Interaction Automated Measures from Combat Vehicle Command and Control Battalion Evaluation*

Instrumented System	Sample of Soldier-Computer Interaction Automated Measures from Combat Vehicle Command and Control Battalion Evaluation
C^4I System	Percent time each map scale used. Percent time each map feature used. Percent time each map scroll function used. Percent control inputs by touch screen. Percent grid inputs to reports by laser device. Number of total reports received*. Number of unique reports received*. Percent and type of reports retrieved from receive queue. Percent and number of reports relayed downward*. Percent and type of reports posted to tactical map. Mean time to retrieve reports. Mean time to relay reports upward and downward*. Number and type of digital reports transmitted*.
CITV System	Percent and number of reports relayed upward*. Percent time in each CITV mode. Number of times laser CITV function used. Number of times Designate target function used.

* Sample measures based on inter-system data examples. All other sample measures in this table are based on intra-system data. CITV = Commander's Independent Thermal Viewer.

Again, from a command and control evaluation perspective, the MWTB infrastructure provides the BCRs an integrated digital data environment. The MWTB environment includes instrumented surrogate C^4I systems coupled with the ability to collect and integrate a wide range of evaluation data with the test bed's DCA system, as indicated in Figure 3. For a more complete description of the BCR program and the MWTB digital infrastructure, see Throne et al. (2000).

The automated measures developed for the BCRs were designed to support its research issues on how digitization might impact future command and staff organizations and functions. In particular, these automated measures focused on the team skills required for effective command and control performance, such as the BCR's streamlined organizations with 13 command and staff participants assigned across four vehicle nodes. These team skill measures were organized under a taxonomy of team processes identified by TRADOC Regulation 350-70 (U.S. Department of the Army, 1999).

The set of automated measures developed for BCR IV are presented in Table 7. The measures are grouped by team process skills. These skill measures address the following team skills: adaptability, shared situational awareness, performance monitoring and feedback, coordination skills, communication skills, and decision-making skills. For a more complete description of the measure development process and an operational definition, rationale, and recommended output formats for each measure, see Throne et al. (2000).

Table 7

Candidate Automated Measures

Team Process Skill Dimension	Candidate Measure of Performance
Adaptability	Terrain Analysis. Node Location. Loss of Node.
Performance Monitoring and Feedback	Situation Report (SITREP) Use. Spot Report (SPOTREP) Use. Commander's Critical Information Requirements (CCIR). Common Map Display. Picture Consistency. Operations Overlay Feedback.
Shared Situational Awareness	Map Area. Sensor Coverage. Satellite Coverage. Line of Sight. Surprise Attack. Collateral Damage. SITREP Lag.
Communication	Whiteboard Use. Radio Communications Pattern. Personnel Initiating Whiteboard Conferences.
Coordination	Overlay Use. Whiteboard Commonality. Targeting. Fire Support Coordination. Fire Engagements. Opposing Forces (OPFOR) Destruction.
Decision-Making	Unmanned Aerial Vehicle (UAV) Effectiveness. Length of Battalion Decision-Making Cycle (Operations Order). Length of Battalion Decision-Making Cycle (Platoon Movement).

Test Planning, Analysis and Evaluation System. A final example of an integrated digital data environment is Test PAES being developed by the U.S. Air Force (Gawron, 2000). As previously noted, the need for more automated measurement methods to reduce the workload associated with more conventional measurement methods is not unique to the Army. As the complexity of Air Force systems and missions increase, test and evaluation becomes more difficult and time consuming. To address this problem, the Air Force developed Test PAES, a computer-based and highly integrated evaluation system that provides for an integrated multimedia data analysis capability. This evaluation system extends the notion of an integrated digital database to include a wide assortment of "cradle-to-grave" databases designed to support evaluation activities ranging from initial formulation of research issues and test plans to final research documentation.

The Test PAES databases include two broad categories of databases: reference and test-specific, as described by Gawron. The reference database includes a library of materials that promote structured and programmatic test procedures. For example, a Card Catalog provides references and abstracts on more than 1,000 technical reports, regulations and test plans as well as templates for more than 50 commonly used report forms. In addition, a database on "best test practices" provides a library of test setups used in former test and evaluation research along with examples of test objectives and guidance on resource requirements. Additional components include standardized training and briefing materials, data analysis plans with sample size requirements, and reporting guidelines with sample data formats. Finally, a lessons learned database describes problems and solutions derived from both Air Force and joint service test and evaluation efforts.

The test-specific databases of Test PAES were designed to help researchers manage information obtained during the conduct of actual test and evaluation exercises. These databases support evaluation functions such as: configuration control of test materials and conditions, event logging to identify actual test events and match their occurrence against planned test events, and a questionnaire database to store and manage data collected by more conventional, manual measurement methods.

The integrative nature of Test PAES, and particularly, its test-specific databases is illustrated in Figure 4. This figure depicts a Data Display Tools menu with data examples taken from a flight test used to evaluate symbology for a head-up display, as described by Gawron. The test point depicted was recovery from an unusual aircraft attitude in which the plane was inverted and nose high, as shown in the side- and rear-view animations titled Flight 1 and Flight 2, respectively. The plots labeled Dynamic 1-3 are time histories with evaluator-selectable parameters, and up to 10 Dynamic windows can be opened concurrently. A timeline of the exercise is provided in the upper right that identifies the start and stop times of this data point. And in the lower right window, video footage of the pilot's view of the head-up display and its symbology for this test point are depicted. Moreover, all of these data display windows are time-synchronized and controlled by the Play Control buttons shown in the "Controls" window.

Data Integration Summary. Although the plethora of data generated by digital systems overwhelms human information processing and integration abilities, digital technologies are increasingly capable of collecting and integrating the data they create. Accordingly, the Army's digitization efforts are increasingly focused on the need to integrate the information required by all users, including evaluators, in the form of a "common" database. Although currently fielded C^4I systems are not fully integrated, the Army's research and development efforts with more advanced C^4I systems provides powerful examples of how digital technologies such as virtual simulation and instrumented C^4I systems can provide an integrated digital data environment for more effective and efficient evaluation methods and measures.

First, the CVCC research program demonstrated the potential of digital technologies to collect and integrate a wide range of performance data, and generate automated measures for command and control training and evaluation. The MWTB's digitally integrated data environment made it possible for the Army to document significant increases in operational effectiveness as a result of digitization at platoon, company, and battalion unit levels (Leibrecht et al., 1994). Second, the BCR research program demonstrated how a digitally integrated data environment supports the skills anticipated from future command and staff organizations, and the evaluation of some of those skills in the form of automated measures of team performance. Finally, the Air Force's Test PAES system exemplified how more extensive database integration can support a cradle-to-grave range of test and evaluation activities.

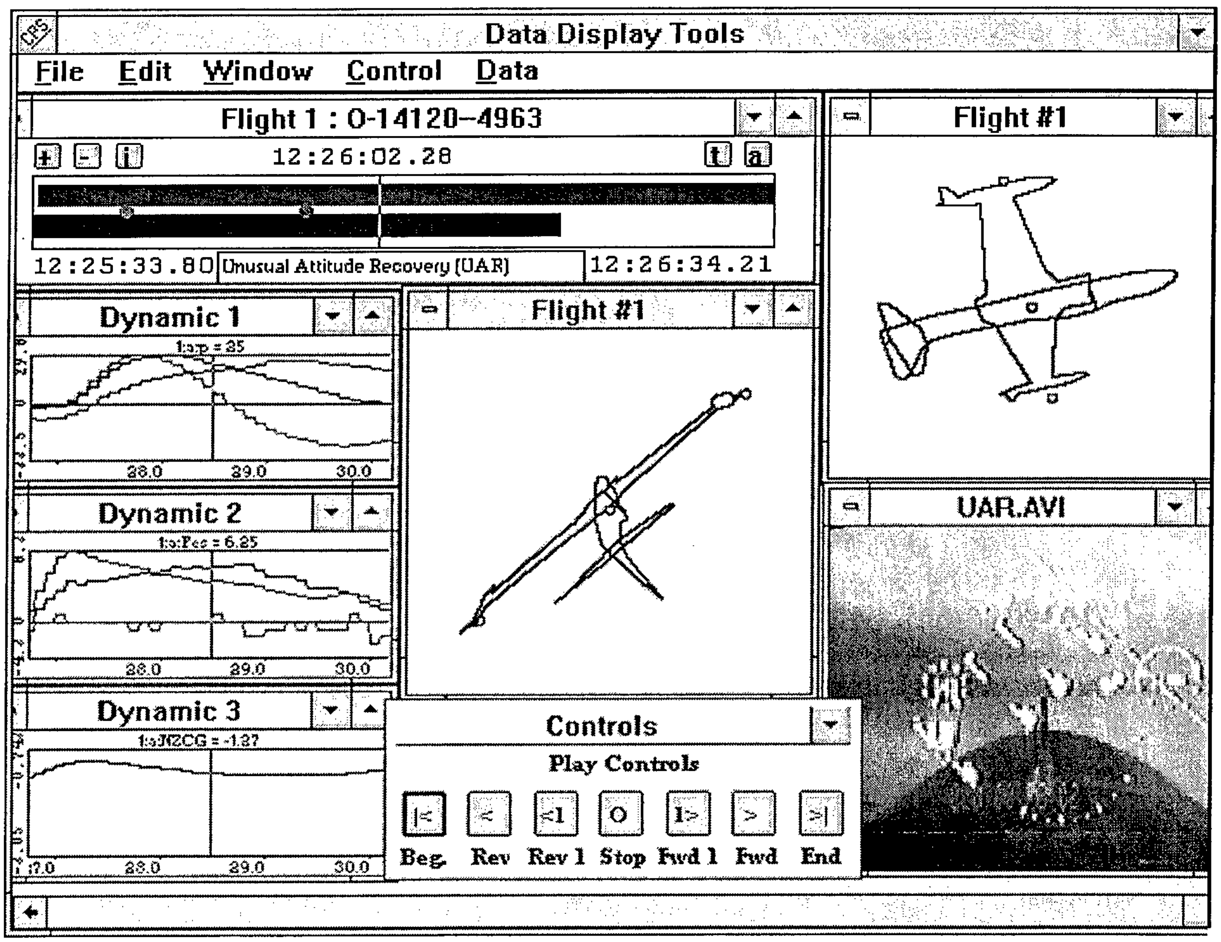


Figure 4. Integrated data display tools provided by the Test Planning, Analysis and Evaluation System. (From "Challenges of Multimedia Design Tools," by V. J. Gawron, 2000, *Ergonomics in Design*, vol. 8, no. 2, p. 28. Copyright 2000 by Human Factors and Ergonomic Society. Reprinted with permission of the author.)

Data Mining

Given the mountains of data often contained in an integrated digital database, data mining tools and techniques are increasingly required to sift through this data to find valuable nuggets of information. The term data mining refers to the process, or set of processes, for sorting through vast amounts of data in order to find useful information. Data mining is a form of information discovery more often associated with exploratory data analysis than confirmatory analysis (Parsaye & Chignell, 1993). Data exploration methods search for patterns and exceptions in data, whereas data confirmation methods confirm whether patterns actually exist. Information discovery is the goal of data mining. Although humans are good at exploration, discovering information in the huge amounts of data compiled in digital databases requires more automated discovery methods. The requirement for digital mining methods in the information age is clear according to Parsaye and Chignell (1993, p. 132) who conclude: "Automatic discovery is the way."

Digital technologies for mining data are not new and are continually improving. Commonly used statistical data exploration tools, for example, include cluster analysis, factor analysis and multidimensional scaling. For a comprehensive review and comparison of data mining tools including: neural nets, machine learning, statistics, rule generation, and anomaly detection, see Parsaye and Chignell (1993). Interested readers might appreciate their rationale for concluding that rule generation and anomaly detection are preferred methods for discovering patterns and relationships within data and information systems, including C^4I systems.

The purpose of this section is to present selected examples of how data mining technology has and might be used to support command and control performance and evaluation. The three data mining examples provided below, range from fundamental techniques for filtering the wealth of digital data generated by training simulations and C^4I systems, to more advanced applications commonly referred to as executive information systems. Each example demonstrates how digital mining tools and techniques can help soldiers as well as evaluators more effectively explore data to discover the information needed in their work. A summary discussion of these data mining examples is provided at the end of this section.

Unit Performance Assessment System. One of the more basic tools for mining data is the use of data filters to extract data elements of interest from a large database. Data filtering requirements are directly related to the volume and complexity of the data generated and available. In simulation-based military training, for example, even smaller scaled exercises at platoon and company level may generate thousands of network data packets each exercise minute (Meliza, Bessemer, & Tan, 1994). The data volume problem is only compounded as exercises expand to larger scales, including multiservice exercises. Although digital tools and methods for filtering data are continually being developed and refined, the following example introduces some basic data filtering issues and approaches.

The Army's development of the Unit Performance Assessment System (UPAS) for networked simulation reinforced the need for data filtering tools and methods (Meliza et al., 1994). Some of the key data filtering requirements identified during UPAS development were the need to limit data collection to one exercise (when multiple exercises were conducted simultaneously), and the need to reduce heavy network data loads. The UPAS documentation underscores the iterative nature of developing effective data filters, as initial approaches may prove either wholly inadequate or only partially successful.

To identify entities unique to a particular exercise, for example, the UPAS developers attempted to filter extraneous data based on a combination of site, host, and entity numbers (e.g., 2.19.03). They quickly discovered that not all entity numbers were available at the start of an exercise, however, and they had to modify the filter by inserting wild card characters (e.g., 2.19.*) to capture data on entities created *during* an exercise, such as semiautomated forces. Often the iterative process of data filter development may require more extensive hardware and/or software upgrades to either the simulation system and/or the data collection system. Software modifications required for UPAS included increasing the size of the data buffer that "holds" protocol data units (PDUs) and revising the program to bundle PDUs into groups of ten and ensure a group versus a single PDU was loaded with each disk access.

To reduce heavy network data loads and prevent the loss of critical data, the UPAS developers explored software modifications designed to selectively eliminate vehicle appearance PDUs. The SIMNET generated various types of PDUs including impact, fire, indirect fire, change in status, and vehicle status as well as vehicle appearance PDUs. However, exercise analysis disclosed that over 90 percent of all PDUs collected were vehicle appearance PDUs. While the loss of other PDU types was considered irreplaceable, the loss of vehicle appearance data was considered to not be a serious problem, as the information from vehicle appearance data from one PDU to the next for the same entity showed little change. Therefore, the UPAS developers modified a data filter so that UPAS would stop collecting vehicle appearance PDUs selectively, after the data buffer was more than 80 percent full.

Critical Information Requirements. A second example of data mining is the use of digital technologies to filter data to satisfy designated critical information requirements. For instance, a commander's critical information requirements (CCIRs) help ensure the commander receives key information and is not overwhelmed with less important information. Typically, humans such as staff personnel, or observers in some training situations, monitor the exercise, filter and synthesize information, and then forward reports to the commander that address the CCIRs. With conventional or analog command and control media the process of meeting CCIRs is "...slow and labor intensive" (Gerber, Henniger & Stone, 1998, p. 1). With digital systems, however, collecting and reporting critical information requirements can be largely automated.

For example, Gerber et al. (1998) describe how a Simulation Information Filtering Tool (SIFT) automatically identifies CCIRs and provides this information to the commander, with the aid of a reporting agent. The SIFT program was initially designed to extract critical information from a constructive simulation called Janus used by the Army and Marine Corps for command and control training. Set up procedures include menu based specification of forces and their activities followed by designation of key terrain areas used in building CCIRs, such as engagement areas, named areas of interest, and phase lines. The CCIRs are then identified via menus used to filter multiple parameters for each of the basic CCIR categories supported by SIFT including "detect and identify," "report movements," "report engagement," "report kills," "report critical combat strength," and "report artillery fire." In addition to more complex parameters for filtering time, radius, detection count thresholds, and combat strength, the basic filter parameters for each type of CCIR include: sides, task forces, equipment, and locations.

After a command and staff training participant builds the desired set of CCIRs, the Janus simulation exercise can begin. Data from the Janus simulation is written to post-processing files, and the SIFT program determines which of these files contain the data needed for each CCIR. For the "report movements" CCIR, for example, the movement data file is opened and read, and the CCIR is processed based upon the parameters specified. When the CCIR parameters are met, the reporting agent sends a CCIR report to the training participant operating a C^4I system called Maneuver Control System Phoenix.

In contrast to the SIFT approach, information filters can also be directly installed into a C^4I system. For example, Force XXI Battle Command Brigade and Below ($FBCB^2$) is a C^4I system with operator-defined "triggers" for: defining time or distance intervals (e.g., updates every five minutes or 500 meters); events (e.g., crossing a phase line), and requests (e.g., tell me where you are right now), as described by Boller (2000).

Similarly, the surrogate C^4I systems used in the previously described Battle Command Reengineering experiments emulate an embedded C^4I routine for meeting CCIRs, including Priority Information Requirements (PIR) about the enemy. Figure 5 presents a sample CCIR/PIR tool menu for the BCR's C^4I system and illustrates a potential user interface for this embedded capability. User procedures for specifying information filters and parameters are similar to those for SIFT, but now performed directly on the C^4I interface. Data from the SIMNET simulation is read directly by the CCIR/PIR application, and the CCIR is processed based upon the parameters specified. When the CCIR parameters are met, the C^4I interface displays an alert. In Figure 5 the user has received a high-level alert (indicated in Red on the user's C^4I display). This alert informs the user that an enemy element of two (2) more vehicles (see Count parameter setting in lower right window) have reached "RoadJ." Moreover, as this tool is embedded in the user's C^4I system, the user can revise or add new information requirements at any time during the simulated exercise.

Executive Information Systems. A final example of how digital technologies support data mining is based on the work of Parsaye and Chignell (1993) who describe the design and use of what are commonly referred to as executive information systems. These systems are powerful software applications designed to meet a wide range of information demands by higher-level workers. Executives and similar leaders are often remote from facts and details. They must rely on inadequate and often unreliable strategies to obtain the voluminous amounts and types of information they need. Data and information glut from digital technologies only exacerbates this dilemma. Of course, the ability of executive information systems to provide information to decision makers requires a well integrated digital database, or more precisely, what the authors call intelligent databases. They define an intelligent database as one that "...manages information in a natural way, making information easy to store, access, and use" (Parsaye & Chignell, 1993, p. 4). An important characteristic of an intelligent database is that it performs tasks decision makers never could in order to aid them, rather than replicate or replace them as with many artificial intelligence applications.

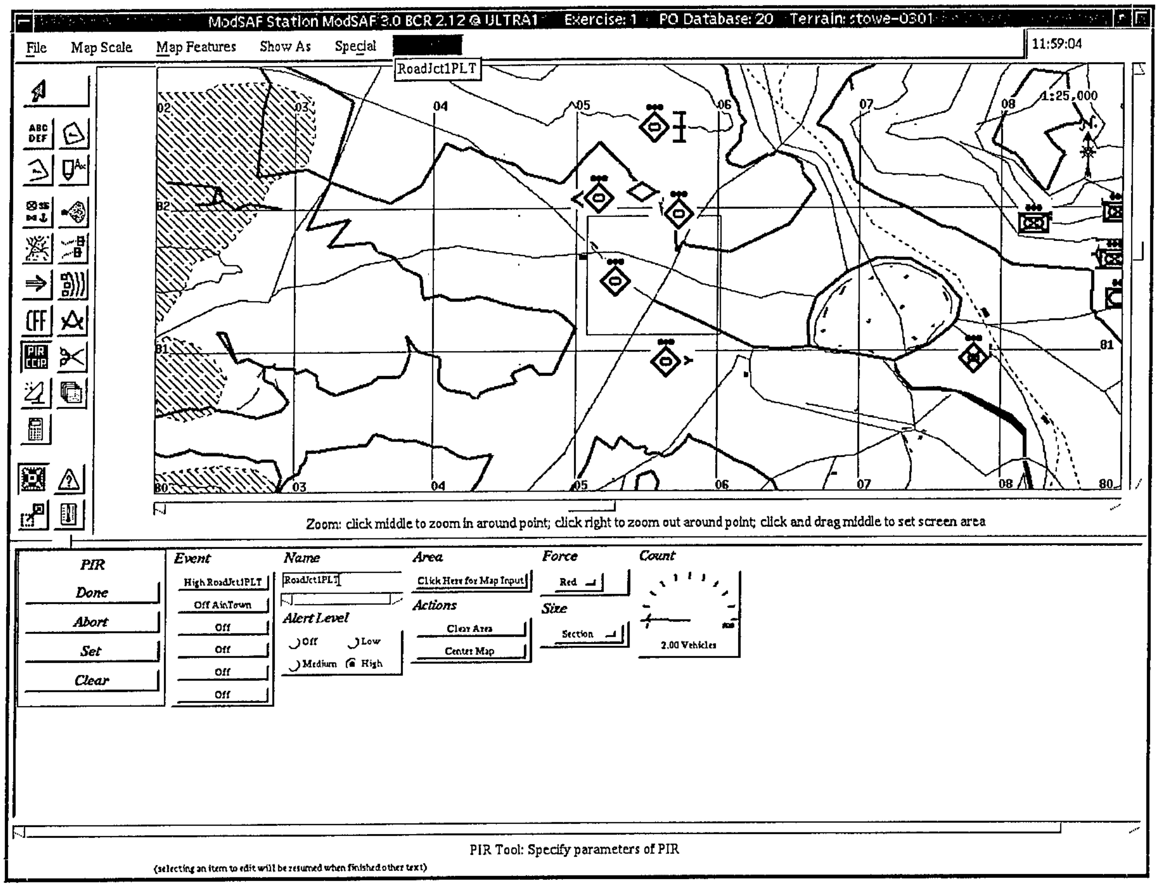


Figure 5. Example of a high-level alert generated by the CCIR/PIR tool on a surrogate C^4I system used for Battle Command Reengineering. (Adapted from U.S. Army Armor Center, 2000).

Parsaye and Chignell (1993) adapt the notion of "the information chain of command" to illustrate how executives depend on information summaries and abstractions prepared by subordinates. They note that this information processing chain applies its own collection and filtering biases that become ingrained and more biasing. As a result, useful information is often lost in the summations provided an executive, or useless due to processing delays required to get information through the chain. Executive information systems resolve this problem in the form of information summaries that provide direct links to underlying data. These links allow executives to "drill down" and review more detailed data, as required. The data mining functions of these systems also filter and sort data in a manner that supports user and organization goals.

A classic design format for an executive information system is an electronic briefing book with point-and-click navigation, suggest Parsaye and Chignell (1993). This output can be tailored to each decision maker's format preferences and areas of concern, including highlighted outcomes, trends, and anomalies. Effective highlights should signify more useful information and alert the user to issues that may be of particular interest. Highlights might prompt the user where to begin a drill down search to data requiring further attention, and also recommend a search path that relates a summary item of concern to its underlying data. The linked data should provide the structure and detail needed for the user to better understand, if not account for, summations and estimations provided. The system should also provide the tools needed to analyze the data in order to explore and identify new relationships and patterns. Finally, to avoid delays and ensure timely information the executive information system should link to *active* databases and the latest information available.

Data Mining Summary. As indicated in this small set of examples, digital technologies are uniquely adept at helping command and control performers and evaluators effectively mine potentially overwhelming amounts of data. Admittedly, many of the UPAS data filtering issues examined might be attributed to earlier, less advanced technology. However, the UPAS example provides detailed documentation on relevant data filtering issues and approaches in the areas of training simulation and the evaluation of command and control performance. As the demand for information and data integration expands, including the military requirement for common databases, data filtering and mining tools must continue their rapid evolvement.

The CCIR/PIR filtering and mining tool examples demonstrate the potential of digital technology to support command and control training and evaluation. They allow commander and staff personnel to repeatedly train and practice with a relatively low-cost, low-manpower simulation. The feedback provided by automated CCIR routines in conjunction with an ongoing simulated exercise should improve the ability of command and control personnel to balance information requirements. A proper balance includes understanding what information to request to ensure requirements are met, as well as what information not to request in order to avoid information overload.

Parallels between executive information systems and C^4I systems, particularly at the commander level, are probably clear to the reader. The growing use of such systems is providing leaders and executives the ability to stay in touch with the facts as well as the people who work with and for them. These systems provide decision makers a bridge between the too often disparate worlds of data and information. Moreover, these systems are electronic support systems that cooperate with and augment the ability of executives, not compete with them. Military commanders, in particular, suffer shortcomings in the "information chain of command" and require the ability to "drill down" to more detailed data, as required. Moreover, the tempo and complexity of future operations require information systems that sort information by priorities, such as the "information triage" categories of dump, delay, and display (Parsaye & Chignell, 1993, p. 328).

Readers concerned about command and control evaluation may have already concluded that many of these data mining examples would improve evaluation measurement methods. Data mining features such as CCIR/PIR tools could be readily adapted to support evaluators' versus performers' critical information requirements. Event based training and evaluation designs, in particular, could develop a set of automated routines that detect and collect key data related to predefined training objectives and/or evaluation measures. At a more global level, evaluators might benefit from the development of an evaluator tool modeled after an executive information system. Clearly, the previously described Test PAES system developed by the Air Force could provide evaluators some of the functionality provided by executive information systems.

Executive information systems appear more capable, however, particularly in their ability to convert data into information, and into output formats that enable data visualization.

Data Visualization

> *With tools for manipulating multimedia information, for arranging it and reconstructing it like the pieces of a puzzle, until a clear image emerges, an individual can be a powerful analyst, and can correctly recreate the many different points of view from which events and ideas might be understood.* (Gano, 1988, p. 261)

Visualization is generally defined as the process of transforming data into visual forms. The range of transformable data includes numbers, objects, and concepts and the range of applications vary from microscopic neural imaging to abstract concept mapping. The purpose of transforming data into visual forms is to enable humans to observe, understand, explore, discover, and experience data in a more meaningful manner. Data visualizations are not about isolated numbers and data points, but about real and potential relationships in numbers (Parsaye & Chignell, 1993). The relations implicit in a database are made explicit in data visualizations that depict patterns, trends, and anomalies.

Digital data visualizations are dynamic and interactive representations. Unlike the static visualizations of analog media, such as print, digital representations enable and promote the ability to query, explore, and reconstruct the information depicted. A simple example is an interactive scatter plot that allows one to click on a depicted outlier and view its source data, or delete a depicted outlier and view a reconstructed scatter plot with the outlier removed. More complex examples of interactive visualization include computer-aided design for development and testing as well as medical imagery used during surgery. In the area of command and control, the interactive microworld visualizations of C^4I systems are expected to support interface inputs and outputs that correspond to movements in the real world.

The importance of visualization tools to command and control appears reflected in Army doctrine that asserts battlefield visualization is critical to mission accomplishment and the art of battle command. Notably, data visualization methods are not equated with the battlefield visualization of commanders or other humans. However, the purpose of developing data visualization methods in digital C^4I systems is to support the ability of humans to visualize the battlefield. Moreover, the Army's definition of battlefield visualization helps clarify how data visualization technologies might improve command and control performance and direct its evaluation.

> *Battlefield visualization is the process whereby the commander develops a clear understanding of his current state with relation to the enemy and the environment, envisions a desired end state and then visualizes the sequence of activities that will move his force from its current state to the end state.* (U.S. Department of the Army, 1997, pp. 1-3)

Similarly, the Army's digitization objective that C^4I systems will provide all combatants and supporters a common picture of the battlefield situation stresses the requirement for data visualization tools and technologies. The requirement for improved visualization tools for command and control performers and evaluators is intensified by the increased complexity and scope of military operations. A key feature of future operations may be an increased reliance on indirect vision, based on emerging trends in information age warfighting such as increased standoff ranges and robotic systems. As line-of-sight operations decrease, digital data visualizations will become soldiers' surrogate "eyes" on the battlefield. To the extent that visualizations replace human vision, the ability to track performers use of battlefield representations will be needed to evaluate command and control performance.

The three examples presented below illustrate how data visualization technology has and might be used to support command and control performance and evaluation. They range from documented examples of how visualization methods support evaluation, to proposed examples of how digital media can transform data into visual forms to improve command and control performance and evaluation. Each example demonstrates how data visualization tools and techniques can help soldiers as well as evaluators more effectively "see" data to discover the information needed in their work. A summary discussion of these data visualization examples is provided at the end of this section.

Macrolevel Visualization. The first example of data visualization is based on the previously described BCR effort to evaluate command and staff processes (Throne et al., 2000). A measure titled Map Area under Shared Situational Awareness (see Table 7) was designed to track participants' use of the BCR's surrogate C^4I systems during test missions. Specifically, Map Area measured the battlefield area visible on each participant's C^4I display at designated times, and tracked the center point of each participant's C^4I display at the same designated times. The four designated times for each mission were the start of the exercise, first direct fire, first indirect fire, and the end of exercise. Visible map area data was provided in tabular format (see Throne et al., 2000), and pictorial format. A sample portion of this BCR's C^4I center-point data in pictorial format is provided in Figure 6.

To provide spatial and performance context, Figure 6 depicts the entire area of the battlefield available to the BCR participants and to their C^4I display system at the time of first indirect fire during a BCR exercise. Also, the squadron's area of operations within this battlefield area, and the relative location and direction of the opposing force (OPFOR) are indicated. Figure 6 plots the center points for each command and staff participant's C^4I display, and indicates that most command and staff members set the center points of their maps in nearly the same location. At first indirect fire, most center points are located near the Eastern edge of the squadron's area of operations (see Figure 6). By the end of this exercise, the immediate threat was defeated and many of the command staff center points had shifted further East toward follow-on enemy forces (see Throne et al., 2000).

Figure 6 also depicts the map area actually *visible* on the C^4I displays of selected participants, namely the squadron commander and a company commander. The instrumentation of these C^4I systems provided the data required to reconstruct visible map area at the times designated. Consider also that visible map area depends on the map scale and zoom levels selected by each participant at any designated time. More importantly, visible map area may reflect the degree to which situational awareness is shared and the informational requirements unique to each duty position. As indicated Figure 6, the squadron commander was maintaining a substantially larger visible map area than the company commander at first indirect fire. The map areas visible to the remaining BCR participants are not depicted in Figure 6 to avoid clutter. However, the commander routinely monitored a map area substantially larger than the area monitored by staff participants, and approximately 2-3 times greater than the squadron's area of operations (see Throne et al., 2000).

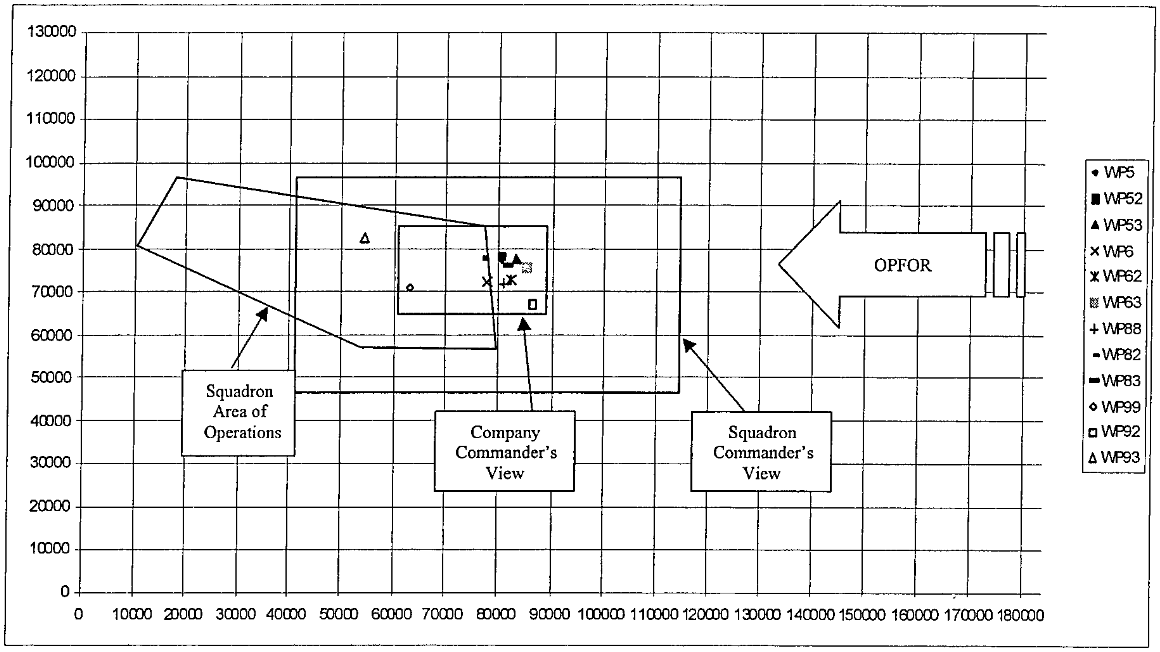


Figure 6. Participants' display center points and visible display areas of squadron commander and company commander, at first indirect fire.

Microlevel Visualization. A quite different example of how digital technologies can help visualize performers' use of situational representations is based on eye track measurements from air traffic controllers. This example comes from research on objective and subjective measurements of human interactions with air traffic management systems and flight deck automation (Jorna, 2000). The evaluators were concerned with how the mental processing limitations of controllers created a bottleneck to expanding the amount of air traffic. Figure 7 depicts a participant controller's eye scanning data, called point-of-gaze transitions, collected with head mounted eye trackers. This data was correlated with a flight simulation exercise to examine how controllers actively used the information on a prototype plan view display (PVD) during low and high traffic loading.

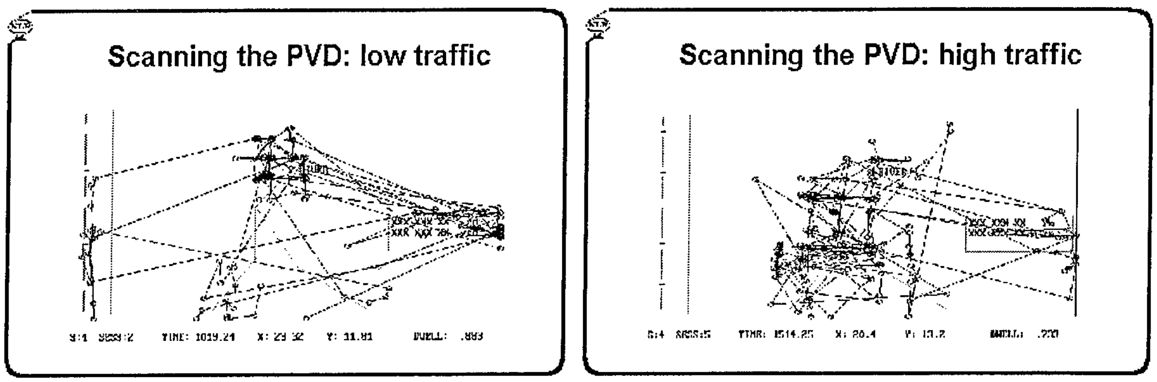


Figure 7. Examples of point of gaze transitions on a plan view display for low (left) and high (right) traffic levels. (From "Human machine interfaces for ATM: Objective and subjective measurements on human interactions with future flight deck and air traffic control systems" by Jorna, 2000, [On-line]. Available: http://vega3.uneec.eurocontrol.fr/jorna.htm Reprinted with permission of the author.)

In turn, Figure 8 illustrates how digital technologies can help evaluators' quantify momentary performance data, such as the eye scan patterns as traced in Figure 7. Figure 8 embeds quantitative performance values of gaze dwell time and fixation frequency data for low and high traffic levels in a spatial configuration of the controller's plan view display. This configuration partitions the display into discrete windows and tools visually accessed by the controller participants including: time line (TL), traffic (TRAF) and handoff (HO) regions. Notably, this visualization example transforms transient performance data into quantitative values directly related to evaluation issues, including human-machine interface design.

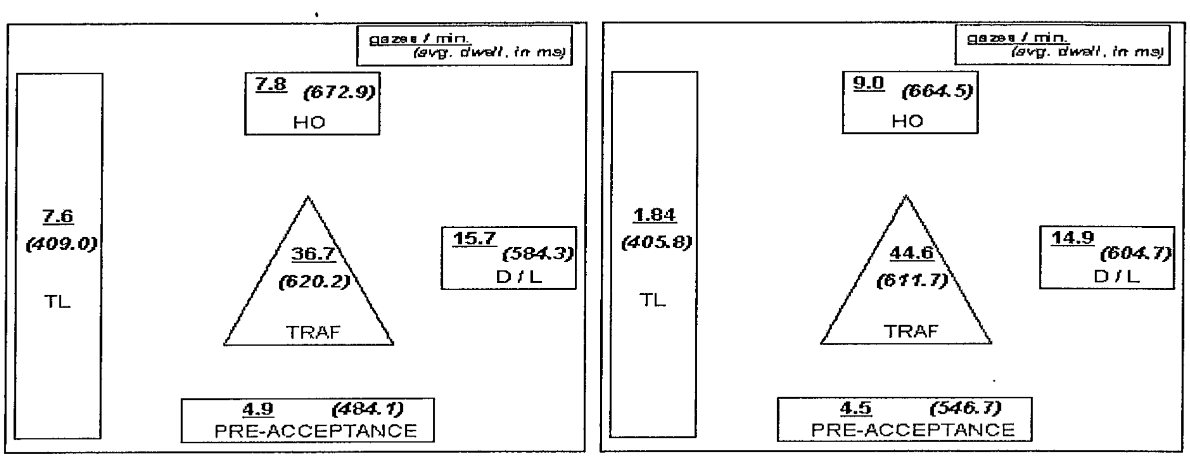


Figure 8. Examples of gaze dwell time and fixation frequencies on a plan view display for low (left) and high (right) traffic levels. (From "Human machine interfaces for ATM: Objective and subjective measurements on human interactions with future flight deck and air traffic control systems" by Jorna, 2000, [On-line]. Available: http://vega3.uneec.eurocontrol. fr/jorna.htm Reprinted with permission of the author.)

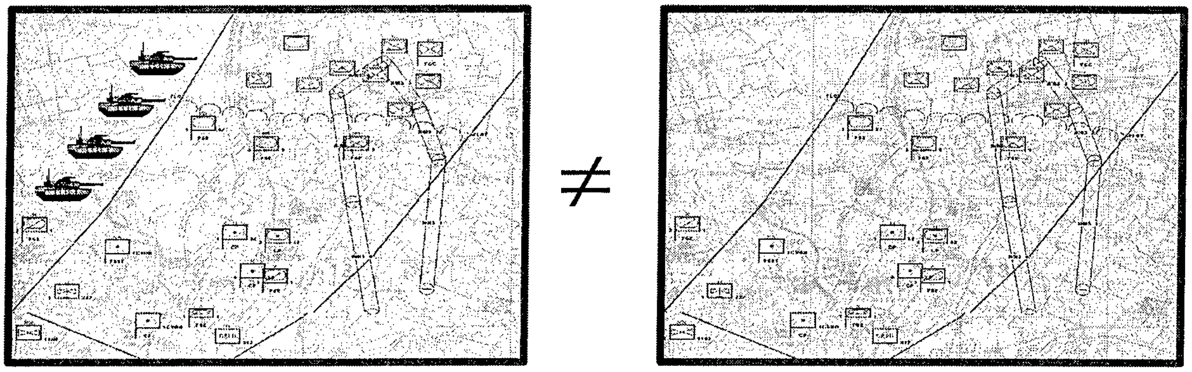
Pictorial Comparison. The final data visualization example is a proposed measurement method that marshals the aid of advanced information systems to improve the match between user and system representations, and their relation to the reality represented. The method, referred to as pictorial comparison, also illustrates how digital evaluation methods might support the Army's objective to provide users a common and accurate picture of battlefield situations. The concept of the common picture provides a meaningful and efficient summation of a complex objective, a communicative power akin to the "same sheet of music" expression. Pictorial comparison based on the information depicted on users' C^4I systems could help transform this concept into an empirically verifiable construct (Lickteig & Throne, 1999).

Pictorial comparison is analogous to the "compare document" function provided by many word processing applications for comparing textual products. For most text applications, this function automatically compares multiple versions of a document and highlights identified discrepancies by source or author. Instrumented C^4I systems could readily provide a similar "compare picture" function. The ability of digital technologies to compare pictorial images is routinely used for fingerprint identification, face recognition, and satellite imagery (Kosinski & Kozlowski, 1998). For example, successive satellite images of clouds are used to measure wind speed, and successive images of ice floes to measure ocean currents in the arctic (Bailey, 1996).

The physical pictures of the battlefield depicted on soldiers' C^4I displays are empirical products composed entirely of pixel data. A display depiction is measurable if it provides access to its underlying database: the data elements depicted on the display and their collaborative source including intra- and inter-system data. This data can be empirically captured in an automated log of soldier-computer interactions. Inter-system data on information received from others would record what battlefield information was available and *not* available to C^4I equipped soldiers, and when that information was available. Intra-system data would disclose when and *if* information was actually visible in the user's display "window" and, to a large extent, what information was examined or ignored by the user.

Figure 9 provides a notional example of how automated pictorial comparison might illustrate a key discrepancy between the battlefield representations depicted on the C^4I displays of two different users at a selected moment during a mission. In this figure, a company commander's display at 1200 hours depicts a platoon unit of four enemy tanks, not simultaneously depicted on the battalion commander's display. Such a potentially important discrepancy would be automatically detected and highlighted by the proposed compare picture method. For other examples, see Brown et al. (1996). Instrumented C^4I systems could extend this example to compare the battlefield situations depicted on the C^4I displays of any or all users, at any or all times, during a simulated mission exercise.

Automated pictorial comparisons are not limited to momentary product comparisons, such as display snapshots. For example, process comparisons could graphically depict information flow, or lack of flow, across the unit. Such a comparison might depict, for example, when and from what source a company commander received the enemy information that was not provided to the battalion commander in Figure 9.



Company Commander's Display
Enemy Platoon NW at 1200 hours

Battalion Commander's Display
No Enemy Platoon NW at 1200 hours

Figure 9. Example of automated pictorial comparison.

Data Visualization Summary. This section examined how data visualization technology has and might be used to support command and control performance and evaluation. The kaleidoscope-like ability of digital technology to iteratively arrange and reconstruct complex data pieces into clear and previously unforeseen patterns and images is a powerful aid to evaluators, as well as performers (Gano, 1988).

The first example presented some relatively macrolevel visualizations of how digital technologies can track and depict the use of C^4I battlefield representations by command and staff participants (see Figure 6). Limitations of this figure include that it is an early prototype in need of refinement and, particularly, the analog media limitations of this report including black-and-white, hard copy. Unlike the static figure depicted here, an interactive visualization would allow evaluators to iteratively rearrange and reconstruct such representations in myriad ways, including the addition of related data such as terrain, obstacles, and selected entity locations. At best, visualizations such as this should exemplify the unprecedented observation and measurement methods that digital technologies afford command and control evaluation. With digital media and methods, evaluators can directly see and examine the battlefield information available to, and made visible by, command and control performers.

The second example provided relatively microlevel visualizations of how digital technologies can track and manifest fine-grained and momentary human-computer interaction data, such as eye scans (see Figures 7 and 8). Moreover, these figures illustrated how such data can be readily quantified and visually related to evaluation issues, such as performers' use of prototype tools during low and high work levels. Two pertinent conclusions from this research that underscore the challenge and opportunity of digital technology are noted. First, when performers are faced with more complex control environments, they may regress to more simple control behaviors. Under low traffic conditions, participants tended to use the tools provided such as time line checks on scheduled arrivals. Under high traffic conditions, however, controllers avoided such tools and reverted to classic 'on screen' controlling methods. Second, digital measurement and visualization methods are needed for evaluating human-computer interactions in complex control environments: "...the use of questionnaires is not sufficient and when not complemented by high quality objective measurements, such a strategy should be qualified as an easy way out" (Jorna, 2000, p. 17).

The final example, pictorial comparison, examined a proposed digital evaluation method for transforming useful concepts into empirical constructs. This method underscores the ability of digital instrumentation to provide a quantifiable, tractable, and visible link between command and control information requirements and systems. Pictorial comparison methods could automatically assess the match between user-system collaborative representations and "ground truth," particularly in digitally integrated simulation exercises.

Pictorial comparisons at a product or process level might be based on normative or criterion standards. Normative comparisons might simply match C^4I pictures across a unit or any designated set of unit members. These comparisons could identify and pictorially depict what is uncommon about a unit's common picture of the battlefield. Although some differences between duty positions are to be expected, the comparisons could be set to address only selected types of information. Only discrepancies in enemy information, for example, are depicted in Figure 9. Other normative comparisons might select other information components, such as mission or terrain, or the commander's critical information requirements.

Criterion comparisons would require an accepted standard of what information should or should not be depicted on a C^4I display. Methods for establishing expert and computer-generated standards for pictorial comparison are described by Lickteig and Throne (1999). Criterion-based comparisons might be used, for example, to train and maintain standard operating procedures for managing user information or maintaining a common picture of the battlefield.

SUMMARY

This report's promotion of digital technologies for evaluating command and control performance begins and ends with the reminder that evaluators "...must use all available weapons of attack..." (Binder, 1964, p. 294). Digital technology represents a new and powerful weapon for attacking evaluation requirements, but is a double-edged sword that poses challenge and opportunity. While digital measurement methods are not a panacea, they can and must help solve many evaluation challenges, including the challenges they create.

The Background chapter acclaims the Army tenet that measurement is integral to training and evaluation, and the Army's tradition of exacting performance standards, proficiency certification, and performance-based training that codify and embody this tenet. However, the Background also notes the Army's renewed requirement for improving evaluation methods, particularly in the area of command and control. Reasons why evaluation concerns are surfacing now include increased complexity and diversity in the contemporary mission environment, personnel turbulence, and the emergence of digital technologies. Evidence reinforcing this requirement includes a pattern of declining readiness at the NTC, the Army's most demanding warfighting arena, attributed in part to a failure to standardize evaluation methods and measures.

The Background's focus on evaluating command and control performance asserts this may constitute the Army's greatest challenge and opportunity in achieving an information age force. A review of behavioral science literature on training evaluation disclosed more problems than answers, and a pattern of evaluation that was too often piecemeal, irrelevant, and subjective. Similarly, a review of conventional command and control evaluation methods disclosed they remain "... sparse, scattered, and of highly variable quality" (Crumley, 1989, p. vii).

The Background closed with a detailed examination of the challenges that confront evaluations of conventional command and control performance. Basic challenges examined include the "art" of command, the indirect relationship of command and control performance to mission outcomes, the nonlinear and context dependent value of information, as well as the complex and collective human-machine information chain that supports command and control. In particular, this analysis focused on two issues: "burdened" methods and measures, and analog media limitations. Evaluations of conventional command and control performance require time consuming and laborious methods and measures that rely heavily on human observation, collection, reduction, integration, analysis, and interpretation. The analog media supporting conventional command and control performance severely limit performers and evaluators.

The Findings chapter examined how the Army might apply digital technologies to evaluation with a focus on command and control performance. This section began with a brief overview of how digital systems are impacting training and evaluation in the Army, and then shifted to the focal issue of digital command and control evaluation. The report's attempt to identify and promote the application of digital technologies to evaluations of command and control performance was divided into three sections: Challenges, Opportunities, and Examples.

Challenges associated with digital command and control evaluation include many of the problems associated with conventional command and control systems, and a host of new challenges. One of the most basic obstacles facing any new system is proving its worth. However, evaluations of units equipped with C^4I systems in live, force-on-force environments have not identified any significant increases in force effectiveness. In addition, the system-of-system complexity problem is currently compounded by introduction and revision of numerous new and largely incompatible C^4I systems.

New challenges include the emerging potential of digital technologies to transform command and control from a force multiplier to a force. For example, the conjunctive nature of digital media couples C^4I systems and weapon systems enabling sensor-to-shooter slew and fire. Similarly, the "productivity paradox" suggests that technology investments require even greater investments in new organizational structures and processes. These "other" investments are costly and take time, and require evaluators employ broader and longitudinal measurement methods. In particular, the pervasive and continuous impact of digital technologies across military structure and practice creates an expanding spiral of evaluation requirements that span the DTLOMS. Evaluation strategies and methods must repeatedly adapt to the ongoing process of change induced by digital technologies.

Opportunities afforded by digital technologies are unprecedented and can help solve many evaluation challenges, including the ones they create. One way of characterizing the opportunity provided by C^4I systems is the expectation they will provide performers and evaluators a common picture of the battlefield situation. Another way is this report's characterization of digital technology as an interactive media for acting on information, and a conjunctive media for accessing information. Digital media not only transmit data to users, they construct information and knowledge *with* and for users. The opportunities afforded by digital technology were organized under the following evaluation issues and sections: Digital Instrumentation, More Balanced Measurement Methods, Increased Scope and Precision, Meaningful Evaluation, and Less Burdened Measurement Methods.

Digital technologies are uniquely suited to automatically collect user performance data. Numerous examples of digital tracking methods and measures were examined ranging from common computer support functions to academic testing and performance evaluation. Despite the demonstrated potential of digital technologies for automatic collection of performance data, that potential is largely unrealized. Most notably, the Army's lack of instrumented C^4I systems limits that potential. As C^4I systems become increasingly integral to the performance of individual and collective tasks, the soldier-computer interactions associated with these systems become more critical and collectible. The report's definition of C^4I instrumentation exploits the conjunctive nature of digital media: instrumentation equates to a log of all soldier-computer interactions correlated with the battlefield situation in which they occur.

Digital evaluation technologies provide direct measures *of* performance. In contrast, a review of training and performance evaluation methods disclosed a biased reliance on measures *about* performance. Measures of performance are needed to balance more conventional observation, interview, and questionnaire methods that often result in overly subjective and imprecise measures. More objective and reliable measures of performance would provide a more solid foundation for evaluating training and mission effectiveness. Clearly, automated measures of performance per se are not a panacea. To correct a perceived imbalance, however, this report focused on more direct measures of performance, or what Prosser called the "dog tracks in the mud" (Prosser, 1964, p. 216).

Digital evaluation technologies should increase measurement precision and scope, particularly in the area of command and control. For a digital unit in simulation-based training, for example, instrumented C^4I systems could provide automated measures of command and control performance for any or all soldiers at any or all times. What types of performance data can be collected with digital technology? This report examined how digital evaluation technologies can assess a wide range of human behavior including psychomotor, procedural, cognitive, and collaborative performance. Cognitive tracking technologies, for example, include the ability to monitor and critique how humans access, filter, compare, integrate, coordinate, and apply information.

Digital media *situate* performers and evaluators in the performance context. This ability supports the need for meaningful measurement methods as performance and evaluation depend on the situation or context. Digital technologies do more than provide compelling performance contexts. They routinely extend and improve the range of performance possible for both performers and evaluators. As C^4I systems become integral to performance, the more they literally situate performers and evaluators in the actual work setting. Accordingly, this report's definition of C^4I instrumentation directly relates performance to context in order to help evaluators understand the purpose of performance, and the adequacy of performance relative to purpose.

Perhaps, the most valuable opportunity provided by digital evaluation technologies is their ability to reduce the measurement burden invariably associated with command and control evaluation. A workload analysis by Brown et al. (1998) forcefully underscored the ability of digital technologies to automate data collection, analysis, and presentation efforts. Overall, they concluded that digital technology could reduce workload on 97 percent of 380 tasks performed by evaluators and analysts. This analysis was based on a notion of C^4I instrumentation limited to *inter*-system data. Instrumentation that includes *intra*-system data, however, might further reduce evaluator workload and improve measurement. Intra-system data, for example, should enable automatic identification and pictorial depiction of what is *uncommon* about a unit's common picture of the battlefield.

Examples of digital measurement methods complete the Findings chapter. These examples illustrate the potential of applying digital technologies to evaluations of command and control. Many of these examples demonstrate the Army's ongoing effort to improve evaluation through the integration of digital technologies, particularly instrumented C^4I systems and virtual simulation. The purported potential of digital measurement methods, however, is not yet realized, even in test beds. The examples were organized under three topics that summarize this potential, and the need for additional research and development to realize that potential: data integration, data mining, and data visualization.

With the advent of digital systems, the world is awash with data that overwhelm human information processing and integration abilities. Fortunately, digital technologies are increasingly capable of collecting and integrating data. This ability is essential to Army digitization objectives that include a common database that provides a common picture of the battlefield. The CVCC research program example demonstrated how digital technologies can collect and integrate a wide range of data across battlefield operating systems, and relate command and control measures of performance to measures of effectiveness. The BCR research program showed how digital data integration supports futuristic command and control skills, and the evaluation of those skills in the form of automated measures of team performance. The Air Force's Test PAES system exemplified how extensive multimedia database integration can support a cradle-to-grave range of test and evaluation activities.

Data mining methods are increasingly required to sift through digital databases and find valuable nuggets of information. The UPAS example documented data filtering issues and approaches related to training simulation as well as command and control evaluation. The CCIR/PIR filtering and mining tool example demonstrated the potential of digital technology to support command and control training and evaluation. For training, these tools allow command and control personnel to repeatedly hone information requirements with low-cost, low-manpower simulation. For evaluation, similar tools can be tailored to meet the information requirements of evaluators. The executive information system example demonstrated how an integrated data collection and analysis system can convert data into output formats that enable data exploration and visualization.

Finally, the report examined how data visualization technology has and might be used to support command and control performance and evaluation. Dynamic and interactive digital data visualizations help evaluators query, explore, and reconstruct the information depicted. The map area example provided macrolevel visualizations that illustrated the use of C^4I battlefield representations by command and staff participants. The eye-tracking example provided microlevel visualizations of transient human-computer interactions, and illustrated how such data can be readily quantified and visually related to evaluation issues. The final example of automated pictorial comparison proposed a digital evaluation method for transforming important concepts into empirical constructs. This example focused on the ability of digital instrumentation to provide a quantifiable, tractable, and visible link between command and control process and products.

Battlefield visualization reflects the Army's digitization objective, to provide a common picture of the battlefield. For command and control performers, data visualization epitomizes the ability of digital technology to foster their battlefield visualization. For command and control evaluators, data visualization epitomizes the ability of digital technology to find "...interocular significance, a result that hits you between the eyes" (Cooley, Gage & Scriven, 1997, p. 20).

CONCLUSIONS

The opportunities afforded by digital technology for improving training and evaluation are as yet only dimly envisioned (e.g., Bailey, 1996; Kurzweil, 1999). The need for improved performance and evaluation is great, however, and such opportunities must not be missed. The intent of this report is to promote the application of digital technologies to command and control performance evaluation. This report began by examining the performance and evaluation challenges confronting conventional and digital command and control. The core of this report examined how digital technologies can and must help solve performance and evaluation challenges, including the ones they create.

Notably, there are technical requirements for improving performance and evaluation, particularly in the area of command and control. Some technical requirements were noted in the Findings chapter including: the need to establish digital connectivity across all C^4I systems, the need to sustain connectivity as technology evolves, and the need to extend connectivity across all military combat, support, and service systems. More specifically, we conclude there are two *critical* technical requirements for applying digital technology to improve command and control performance and evaluation. These two requirements, and the need for *standards* to ensure these requirements are met, are bulleted and briefly discussed.

- Require instrumentation of all C^4I systems, particularly training and operational systems.
- Establish compelling standards to assess this C^4I instrumentation requirement.
- Require technology integration, particularly C^4I systems and training simulations.
- Establish compelling standards to assess this technology integration requirement.

A critical requirement for achieving the evaluative potential of digital technologies is the instrumentation of these technologies, particularly C^4I systems. As C^4I systems become integral to the performance of individual and collective tasks, the tractable human-computer interactions associated with these systems become more critical and collectable. The more a C^4I interface becomes the primary means of interaction between soldiers and the battlefield, the more instrumented data on human-computer interactions represent essential process and product metrics for command and control evaluation. Standards to assess this requirement can be derived from our proposed definition: *instrumentation of C^4I systems equates to a log of all soldier-computer interactions correlated with the battlefield situation in which they occur.* These standards should reinforce the need for *inter-* and *intra-*system data to improve evaluation and reduce evaluation workload.

Another critical requirement is full integration of digital technologies, particularly C^4I systems and training simulations. Integrated technologies are needed to realize the conjunctive potential of digital media and Army objectives. Digitally integrated technologies afford a unique opportunity to directly *situate* performers and evaluators in the performance context. Digital representations of contextual conditions extend and improve the range of performance possible, and provide an experiential and empirical basis for meaningful evaluation. Standards to assess technology integration should precisely *correlate* task conditions, performance, and performance measures. Integration standards should reinforce the Army's overarching digital objectives, such as the Army Common Database and a common picture of the battlefield for all combatants and supporters.

After decades of developing and fielding C^4I systems, limitations in system integration and instrumentation persist and severely impact the Army's performance, training and evaluation capabilities. This report concludes that system integration and instrumentation requirements are critical to force development objectives. Integration and instrumentation requirements must be identified and articulated in ongoing and future force development efforts, including Future Combat Systems and Objective Force requirements (Cohen, 2000). All developmental efforts related to digital technology including Operational Requirements Documents (ORDs) for new and current systems should define and enforce system integration and instrumentation requirements. In sum, system integration and instrumentation are *prerequisite* requirements for realizing the opportunities afforded by digital technology for improving performance and evaluation, particularly in the area of command and control.

Opportunity entails effort. The conventional refrain "more research and development is required" applies in spades for digital. While this report underscores the need for a digital evaluation strategy, it is not a blueprint or a handbook of digital measurement methods. Much work remains. However, the opportunities examined here rest firmly on empirical methods for automating measures of performance, for increasing the scope, precision and meaning of measures, and for decreasing the burden of measurement and evaluation.

Clearly, the work that remains in applying digital evaluation methods is substantial and requires concerted effort by Army research, development, training, and evaluation personnel. Moreover, this work requires a sustained effort as evaluation strategies and methods must adapt repeatedly to the ongoing process of change induced by digital technologies. As evaluators learn by applying digital technology and as the technology advances, digital evaluation and measurement methods will evolve far beyond the limited examples examined in this report. Learning is in the doing, and as Binder (1964) admonished, we must first "learn to sweat."

References

Ackerman, P. L., Cianciolo, A. T. (1999). Psychomotor Abilities Via Touch-Panel Testing: Measurement Innovations, Construct, and Criterion Validity. *Human Performance, 12,* ¾, 231-273.

Alberts, D. S., Garstka, J. J., & Stein, F. P. (2000). *Network centric warfare: developing and leveraging information superiority, 2nd edition* (revised) [On-line]. Available: http://www.ida.org/DIVISIONS/sctr/cpof/2ndNCW.pdf

Allen, P. D. (1999). Don't hide your assumptions: A C^4I analysis approach that focuses on the assumptions as critical to the cause-and-effect trace [On-line]. Available: http://www.mors.org/c4isr/c4i.htm

Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. *Human Factors, 33,* 343-362.

Bailey, J. (1996). *After thought: The computer challenge to human intelligence.* New York, NY: Basic Books, A Division of HarperCollins Publishers.

Bennet, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. *Educational Measurement Issues and Practice, 17* (4), 9-16.

Binder, A. (1964). Statistical Theory. In P. R. Farnsworth, O. McNemar, & Q. McNemar (Eds.), *Annual Review of Psychology, 15,* (277-310).

Boller, M. (2000). A common understanding for transformation brigades. *Military Review,* September-October, 29-38.

Bornman, L. G. (1993). *Command and Control Measures of Effectiveness Handbook.* (TRADOC Technical Document, TRAC-TD-0393). Fort Monroe, VA: Author.

Brown, F. (1999, June 4). *Army learning white paper: Leader preparation* (Draft). Alexandria, VA: Institute for the Defense Analyses.

Brown, B., Nordyke, J. Gerlock, D., Begley, I. & Meliza, L. (1998). *Training analysis and feedback aids (TAAF Aids) study for live training support* (ARI Study Report 98-04). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Brown, B. R., Nordyke, J. W., Gerlock, D. L., Begley, I. J. II, & Meliza, L. L. (1998). *Training Analysis and Feedback Aids Study for Live Training Support* (ARI Study Report 98-04). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Brown, B., Metzler, T., Riede, D., & Wonsewitz, R. (1996). *Exploiting artificial intelligence to improve after action reviews for the digitized company team.* (Report No. 1157-0002AB). Lawton, OK: Advancia.

Brynjolfsson, E. & Hitt, L.M., (1998). Beyond the productivity paradox: Computers are the catalyst for bigger changes, *Communications of the ACM, 41,* (8) 49-55.

Burroughs, W. A., Murray, J., Wesley, S. S., Medina, D. R., Penn, S. L., Gordon, S. R., Catello, M. (1999). Easing the Implementation of Behavioral Testing Through Computerization. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations In Computerized Assessment* pp. 221-248. New Jersey: Lawrence Earlbaum Associates.

Cannon-Bowers, J. A., & Salas, E. (1997). A framework for developing team performance measures in training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement* pp. 45-62. Mahwah, NJ: Lawrence Erlbaum Associates.

Chignell, M. & Waterworth, J. (1997). Multimedia In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1808-1861). New York: John Wiley and Sons, Inc.

Cohen, W. S. (Secretary of Defense) (2000). *Annual report to the president and the congress* [On-line]. Chapter 11, A Strategy for Military Transformation, Available: http://www.dtic.mil/execsec/ adr2000/chap11.htm

Cook, R. I., & Woods, D. D. (1996). Adapting to new technology in the operating room. *Human Factors, 38* (4), 593-613.

Cooley, W. W., Gage, N. L. & Scriven, M. (1997). The Vision Thing: Educational Research and AERA in the 21st Century Part 1: Competing Visions of What educational Researchers Should Do. *Educational Researcher, 26*, 4, 18-21.

Crumley, L. M. (1989). *Review of research and methodologies relevant to Army command and control performance measurement* (ARI Technical Report 825). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Decker, G. F. (1996). The U.S. Army towards Force XXI. *Military Technology, 20,* 111-117. (ISSN 0722-3226)

Dierksmeier, F. E., Johnston, J. C., Winsch, B. J., Leibrecht, B. C., Sawyer, A. R., Quinkert, K. A. & Wilkinson, J. G. (1999). *Structured simulation-based training program for a digitized force: approach, design, and functional requirements, Volume I* (ARI Research Report 1737). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

DiSessa, A. (1986). Artificial worlds and real experience. *Instructional Science, 14*, 207-227.

Ehret, B. D., Gray, W. D., & Kirschenbaum, S. S. (2000). Contending with complexity: Developing and using a scaled world in applied cognitive research. *Human Factors, 42*, (1), 8-23.

Ensing, A. R. & Knapp, B. G. (1995). *Modeling army C2 at the brigade level.* Paper presented at 1995 International Symposium of Command and Control Research and Technology, Washington, DC.

Finley, D. L., Muckler, F. A., Gainer, C. A., & Obermayer, R. W. (1975). *An analysis and evaluation methodology for command and control: Final technical report.* Arlington, VA: Naval Analysis Programs, Office of Naval Research.

Fletcher, J. D. (1994). What network simulation offers to the assessment of collectives. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in software applications* pp. 255-272. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gano, S. (1988). Multimedia Technology is for casual, everyday use. In S. Ambron & K. Hooper, (Eds.), *Interactive multimedia: Visions of multimedia for developers, educators and information providers* pp. 253-265. Redmond, Washington: Microsoft Press.

Gates, B. (1996). *The Road Ahead.* New York, NY: Penguin Books.

Gawron, V. (2000). Challenges of multimedia tool design. *Ergonomics in Design,* Spring 2000, 22-28.

General Accounting Office (1999). Military Readiness: *Full training benefits from army's combat training centers are not being realized* (GAO/NSIAD-99-210). National Security and International Affairs Division.

Gerber, W. J., Henninger, A. J. & Stone, G. F. (1998). Intelligent extraction, filtering and reporting of critical decision information from a simulation. *Proceedings of the Simulation Interoperability Workshop, International Ergonomics 2000/Human Factors Ergonomics Society 2000 Congress*, I-303-I-307.

Gerhardt-Powals, J. (1996). Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction, 8,* (2), 189-211.

Google. (2000). [On-line]. Available: http://www.google.com/search, 6 Nov 2000.

Guerlain, S. A., Smith, P. J., Obradovich, J. H., Rudman, S., Strohm, P., Smith, J. W., Svirbely, J., & Sachs, L. (1999). Interactive critiquing as a form of decision support: an empirical evaluation. *Human Factors, 41*, (1), 72-89.

Guzidal, M. & Turns, J. (2000). Computer-supported collaborative learning in engineering: the challenge of scaling-up assessment. In M. J., Jacobson, & R. B. Kozma, (Eds.). In *Innovations in science and mathematics education: advanced designs for technologies of learning* pp. 227-257. New Jersey: Lawrence Earlbaum Associates.

Hallmark, B. W., Crowley, J. C., Leonard, H., Lippiatt, T. F. & Sollinger, J. (1998). *Determining training proficiency at combat training centers: data collection instruments* (RAND DRR-1954-A). Santa Monica, CA: Arroyo Center.

Halpin, S. (1992). *Army Command and Control Evaluation System (ACCES): A Brief Description.* (ARI Research Note LVN 92-01). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Helms, J., Neale, D., Isenhour, P. & Carroll, J. (2000). Data logging: Higher-level capturing and multi-level abstracting of user activities. *Proceedings of the International Ergonomics 2000/Human Factors Ergonomics Society 2000 Congress*, I-303-I-307.

Herl, B. (1999). The graphical spot report: A multiechelon technique to quickly track and record battlefield information. *Armor*, September-October, 51-52.

Hooper, K. (1988). Multimedia in education. In S. Ambron & K. Hooper, (Eds.), *Interactive multimedia: Visions of multimedia for developers, educators and information providers* pp. 315-330. Redmond, Washington: Microsoft Press.

Jorna, P. G. A. M. (2000). *Human machine interfaces for ATM: Objective and subjective measurements on human interactions with future flight deck and air traffic control systems* [On-line]. Available: http://vega3.uneec.eurocontrol.fr/jorna.htm

Kass, R. (1996). *Methodology for operational tests and experiments* (Draft 15 May 1996). Advanced Concepts Test and Integration Directorate, TEXCOM.

Kirlik, A. & Bisantz, A. (1999). Cognition in human-machine systems: Experiential and environmental aspects of adaptation. In P. Hancock, (Ed.), *Human Performance and Ergonomics* pp. 47-68. San Diego, CA: Academic Press.

Kittler, F. (2000). *The history of communication media.* [On-line]. Available: http://www.ctheory.com/ga1.14.html

Kontogiannis, T. (1999). Applying information technology to the presentation of emergency operating procedures: Implications for usability criteria. *Behavior and Information Technology, 18*, (4) 261-276.

Kosinski, R. & Kozlowski, C. (1998). Artificial neural networks—modern systems for safety control. *Safety and Ergonomics, 4*, (3), 317-332.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61,* (2), 179-211.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42,* (2), 7-19.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence.* New York, NY: Penquin Books.

Leibrecht, B. C., Meade, G. A., Schmidt, J. H., Doherty, W. J., & Lickteig, C. W. (1994). *Evaluation of the combat vehicle command and control system: Operational effectiveness of an armor battalion* (ARI Technical Report 998). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lickteig, C. W. & Throne, M. H. (1999). *Applying digital technologies to training: A focus on pictorial communication* (ARI Technical Report 1097). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Locatis, C. (2000). *An interactive technology primer.* [On-line]. Available: http://tlc.nlm.nih.gov/resources/publications/primer/coverpage.html

Meliza, L. L., Bessemer, D. W., & Tan, S. C. (1994). *Unit performance assessment system (UPAS) development* (ARI Technical Report 1008). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Metersky, M. L. (1986). A C2 process and approach to design and evaluation. *IEEE Transactions on Systems, Man, and Cybernetics*, (Special Issue on Information Technology for Command and Control) *SMC-16*, (6), 880-890.

McLuhan, M. (1964). *Understanding media: the extensions of man.* New York, NY: McGraw-Hill Book Company.

McLuhan, M., Fiore, Q. & Agel, J. (1967). *The medium is the massage.* New York, NY: Random House Inc.

Moses, F. (in preparation). *Training research plan to address army battlefield digitization* (ARI Special Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Nickerson, R. S., (Ed.). (1995). *Emerging needs and opportunities for human factors research.* Committee on Human Factors, National Research Council. Washington, DC: National Academy Press.

Nolan, J. M. (1999). Using flag signals to speed combat casualty evacuation. *Armor*, September-October, 39-40.

Office of Management and Budget (1993). *Government performance results act of 1993.* [On-line]. Available: http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html

Olmstead, J. A., Christensen, H. E., & Lackey, L. L. (1973). *Components of organizational competence: test of a conceptual framework* (HumRRO TR-73-19). Alexandria, VA: Human Resources Research Organization.

Olmstead, J. A., Elder, B. L., & Forsyth, J. M. (1978). *Organizational process and combat readiness: Feasibility of training organizational effectiveness staff officers to assess command group performance.* (ARI Technical Report 468). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Parsaye K. & Chignell, M. (1993). *Intelligent database tools and applications: Hyper-information access, data quality, visualization, automatic discovery.* New York, NY: John Wiley & Sons, Inc.

Per-Arne Persson (1999). Transformation of the analog: The case of the Saab BT 33 artillery fire control simulator and the introduction of the digital computer as control technology. *IEEE Annals of the History of Computing, 21*, (2), 52-63.

Pejtersen, A. M., & Rasmussen, J. (1997). Effectiveness testing of complex systems. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* 2nd ed., pp. 1514-1542. New York: John Wiley & Sons, Inc.

Phelps, R. O. & Kupets, G. (1984). *U.S. Army Human Engineering Laboratory Communications Survey—A pilot study (HELCOMS_PS)* (Technical Note 14-84). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

Prosser, W. L. (1964). *Handbook of the law of torts.* (3rd edition) St. Paul: West.

Rasmussen, J., & Pejtersen, A. M. (1995). Virtual ecology of work. In J. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.), *Global perspectives on the ecology of human-machine systems* pp. 121-156. Hove, UK: Lawrence Earlbaum Associates.

Reidel, S. L. & Fitz, G. F. (1986). Utilization-oriented evaluation of decisions support systems. *IEEE Transactions on Systems, Man, and Cybernetics*, (Special Issue on Information Technology for Command and Control) *SMC-16*, (6), 980-996.

Rosenberger, J. D. (1999). Reaching our Army's full combat potential in the 21st century. *Armor, 46,* May-June, 8-14.

Salas, E., Rhodenizer, L., & Bowers, C. A. (2000). The design and delivery of crew resource management training: exploiting available resources. *Human Factors, 42*, (3), 490-511.

Sands, W. A, Waters, B. K. & McBride, J. R. (1999). CATBOOK computerized adaptive testing: from inquiry to operation (ARI Research Note 99-02). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sarter, N. B., Woods, D. D. & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* 2nd ed., pp. 1926-1943. New York: John Wiley and Sons, Inc.

Schorow, S. (2000). Spinning out of control? While society embraces info-technology, some see a world gone hay-wired. [On-line]. Available: http://www.bostonherald.com/lifestyle/lifestyle_trends/fast10242000.htm

Shields, J. L., Cavallaro, J. B., Huey, B. M. & Van Cott, H. P. (1995). Training and education. In Nickerson, R. S., (Ed.) *Emerging Needs and Opportunities for Human Factors Research.* Committee on Human Factors, National Research Council. Washington, DC: National Academy Press, 86-105.

Throne, M. H., Deatz, R. C., Holden, W. T. Jr., Campbell, C. H., Sterling, B. S., & Lickteig, C. W. (1999). *Prototype staff training and evaluation methods for future forces* (ARI Research Report 1745). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Throne, M. H., Holden, W. T. Jr., Lickteig, C. W. (2000). *Refinement of prototype staff evaluation methods for future forces: a focus on automated measures* (ARI Research Report 1764). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

U.S. Army Armor Center (2000). *User's manual for battle command reengineering.* Available: Commander, U.S. Army Armor Center, ATTN: ATZK-MWT, Fort Knox, KY 40121-5000.

U.S. Army Training and Doctrine Command (1995). *Training development management, processes, and products* (TRADOC Regulation 350-70). Fort Monroe, VA: Author.

U.S. Army Training Support Center (1996). *Report on live environment research requirements.* Available from the U.S. Army Training Support Center, Fort Eustis, VA.

U.S. Department of the Army (1995). *Force XXI Operations: A concept for the evolution of full-dimensional operations for the strategic army of the early twenty-first century.* (TRADOC Pamphlet 525-5). Fort Monroe, VA: Author.

U.S. Department of the Army (1988). *Field Manual 25-100, Training the force.* Fort Leavenworth, KS: Combined Arms Center.

U.S. Department of the Army (1996). *Training guide for producing collective training products* (TRADOC Pam 350-70-1). Fort Monroe, VA: Deputy Chief of Staff for Training.

U.S. Department of the Army (1997). *Field Manual 101-5, Staff Organization and Operations.* Washington, DC: Headquarters Department of the Army.

U.S. Department of the Army (1998a, September). *TRADOC digital learning strategy* (Memorandum). Fort Monroe, VA: Deputy Chief of Staff for Training.

U.S. Department of the Army (1998b, October). *Information management: automation TRADOC plan for reengineering information systems modernization (TPRISM)* (TRADOC Pam 25-73). Fort Monroe, VA: Deputy Chief of Staff for Training.

U.S. Department of the Army (1999). *TRADOC regulation 350-70, systems approach to training management, processes, and products*. Fort Monroe, VA: Headquarters, U.S. Army Training and Doctrine Command.

Ward, J. H. (2000). A historical perspective...the U. S. army information systems engineering command's technology integration center. *Army Acquisition Logistics and Technology*, Nov-Dec 2000, 44-45.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest. L. (1973). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally & Company: Chicago.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mustapha (Eds.), *Automation and human performance: theory and applications* pp. 3-17. New Jersey: Lawrence Earlbaum Associates.

Appendix A

Acronyms

AAR	After Action Review
ACCES	Army Command and Control Evaluation System
ARI	Army Research Institute
ARTEP	Army Training and Evaluation Program
ASAT	Automated Systems Approach to Training
AWE	Advanced Warfighting Experiment
BCR	Battle Command Reengineering
BLUFOR	Blue Forces
BOS	Battlefield Operating System
C^4I	Command, Control, Communication, Computer, and Intelligence
C^4ISR	Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance
CALL	Center for Army Lessons Learned
CCIR	Commander's Critical Information Requirement
CCTT-D	Close Combat Tactical Trainer - Digital
CEP	Concept Experimentation Program
CITV	Commander's Independent Thermal Viewer
COA	Course of Action
CTC	Combat Training Center
CVCC	Combat Vehicle Command and Control
DCA	Data Collection Analysis
DTLOMS	Doctrine, Training, Leadership, Organization, Materiel, and Soldier
FBC	Future Battlefield Conditions
$FBCB^2$	Force XXI Battle Command Brigade and Below
FRAGO	Fragmentary Order
FUTURE-TRAIN	Techniques and Tools for Embedded Collective Training of Future Brigade Combat Team Commanders and Staffs
GAO	General Accounting Office
HO	Handoff
INTEL	Intelligence Report
MMBL	Mounted Maneuver Battle Lab
MOA	Memorandum of Agreement
MOE	Measure of Effectiveness
MOP	Measure of Performance

MTP	Mission Training Plan
MWTB	Mounted Warfare Test Bed
NTC	National Training Center
OC	Observers/Controllers
ORD	Operational Requirements Document
OPFOR	Opposing Force
OPORD	Operations Order
PAES	Planning, Analysis and Evaluation System
PDU	Protocol Data Unit
PIR	Priority Information Requirement
POI	Program of Instruction
PVD	Plan View Display
SIMNET	Simulation Networking
SIFT	Simulation Information Filtering Tool
SITREP	Situation Report
SOP	Standard Operating Procedure
SPOTREP	Sport Report
TAF	Training Analysis Facility
T&EO	Training Evaluation Outline
THP	Take Home Package
TL	Time Line
TOC	Tactical Operations Center
TRADOC	Training and Doctrine Command
TRAF	Traffic
TSP	Training Support Package
UAV	Unmanned Aerial Vehicle
UPAS	Unit Performance Assessment System
USAARMC	U.S. Army Armor Center